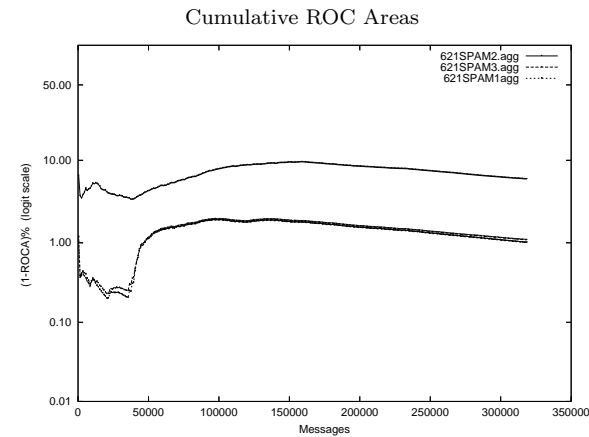
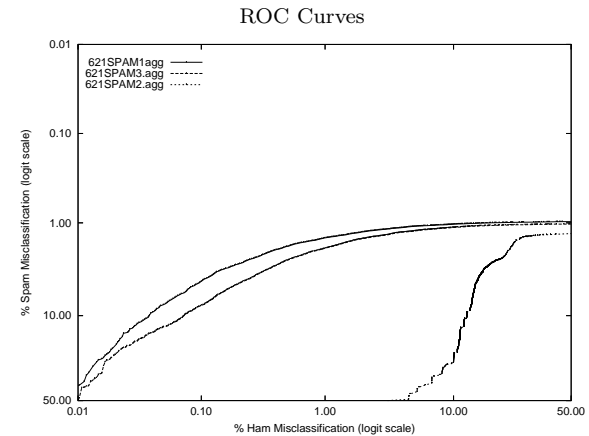
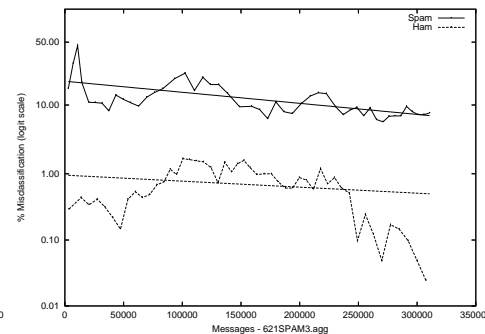
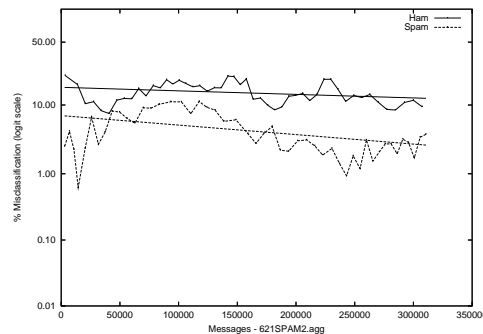
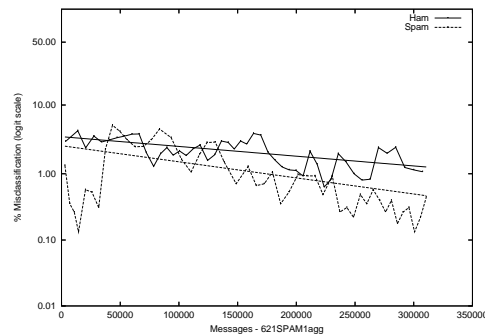


621 - IBM Research (Segal) on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	621SPAM1agg			621SPAM2.agg			621SPAM3.agg		
	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	spam
Ham Misc%	2.20 (2.14-2.26)			14.59 (14.43-14.74)			0.70 (0.66-0.74)		
Spam Misc%	1.23 (1.17-1.30)			4.50 (4.38-4.62)			12.58 (12.39-12.77)		
Logit av Misc%	1.65 (1.60-1.70)			8.23 (8.12-8.35)			3.08 (3.00-3.17)		
(1-ROCA)%	1.008 (0.960 - 1.058)			6.064 (5.968 - 6.162)			1.090 (1.033 - 1.150)		
Ham Misc%	Spam Misc%								
0.01	40.67 (27.39 - 55.47)			54.35 (53.50 - 55.21)			51.31 (32.72 - 69.54)		
0.02	20.38 (14.07 - 28.59)			54.34 (54.08 - 54.60)			23.20 (16.17 - 32.13)		
0.05	7.69 (6.54 - 9.02)			54.29 (54.03 - 54.55)			12.86 (11.53 - 14.32)		
0.10	4.36 (3.91 - 4.87)			54.21 (53.95 - 54.46)			7.89 (7.19 - 8.65)		
0.20	2.84 (2.70 - 2.98)			53.98 (53.72 - 54.23)			4.66 (4.33 - 5.02)		
0.50	1.82 (1.74 - 1.90)			53.45 (53.19 - 53.71)			2.58 (2.45 - 2.72)		
1.00	1.47 (1.40 - 1.53)			52.50 (52.23 - 52.77)			1.91 (1.83 - 2.00)		
2.00	1.26 (1.20 - 1.32)			51.18 (50.91 - 51.46)			1.50 (1.43 - 1.56)		
5.00	1.10 (1.04 - 1.15)			44.98 (44.60 - 45.36)			1.24 (1.17 - 1.30)		
10.00	1.03 (0.98 - 1.09)			27.12 (24.28 - 30.15)			1.12 (1.06 - 1.19)		
20.00	0.99 (0.95 - 1.05)			2.60 (2.49 - 2.70)			1.07 (1.01 - 1.13)		
50.00	0.97 (0.92 - 1.02)			1.31 (1.24 - 1.38)			1.03 (0.97 - 1.09)		
Spam Misc%	Ham Misc%								
0.01	100.00 (100.00 - 100.00)			100.00 (nan - nan)			100.00 (99.99 - 100.00)		
0.02	100.00 (99.99 - 100.00)			99.99 (99.99 - 100.00)			99.99 (99.98 - 100.00)		
0.05	99.98 (99.97 - 99.99)			99.99 (99.98 - 99.99)			99.99 (99.97 - 99.99)		
0.10	99.96 (99.95 - 99.97)			99.96 (99.95 - 99.97)			99.95 (99.93 - 99.96)		
0.20	99.93 (99.92 - 99.95)			99.92 (99.90 - 99.94)			99.90 (99.88 - 99.92)		
0.50	99.83 (99.81 - 99.85)			99.83 (99.80 - 99.85)			99.78 (99.75 - 99.81)		
1.00	17.67 (1.68 - 72.98)			95.98 (84.47 - 99.05)			99.58 (57.81 - 100.00)		
2.00	0.39 (0.35 - 0.44)			23.91 (23.30 - 24.53)			0.88 (0.78 - 1.00)		
5.00	0.08 (0.07 - 0.10)			14.22 (14.05 - 14.39)			0.18 (0.16 - 0.20)		
10.00	0.04 (0.03 - 0.05)			12.46 (12.31 - 12.60)			0.07 (0.06 - 0.08)		
20.00	0.02 (0.02 - 0.03)			10.82 (10.65 - 11.00)			0.02 (0.02 - 0.03)		
50.00	0.01 (0.00 - 0.01)			3.53 (3.17 - 3.93)			0.01 (0.01 - 0.02)		

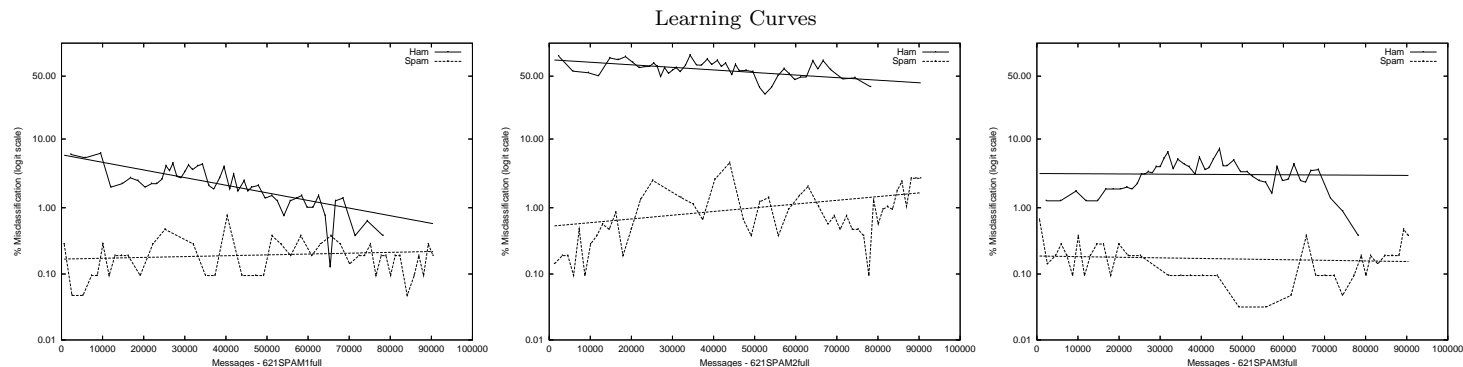
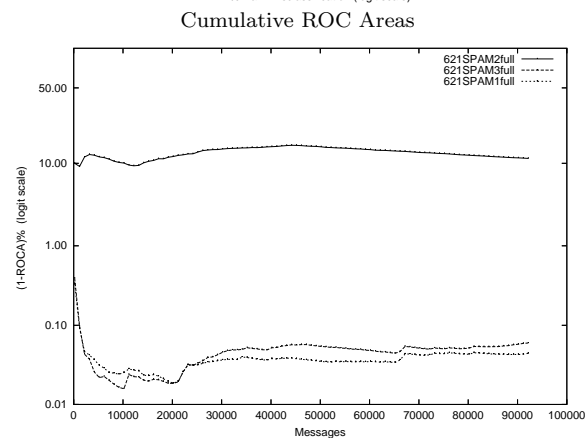
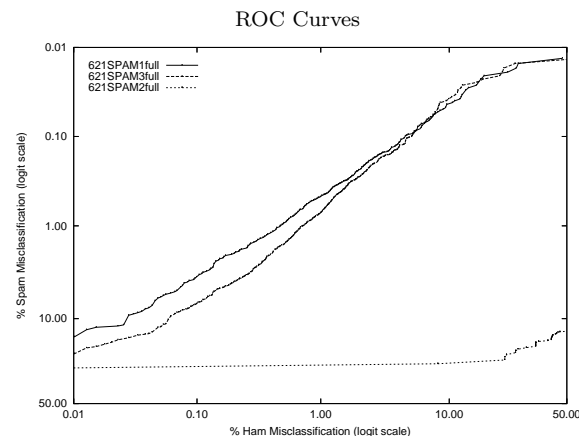


Learning Curves



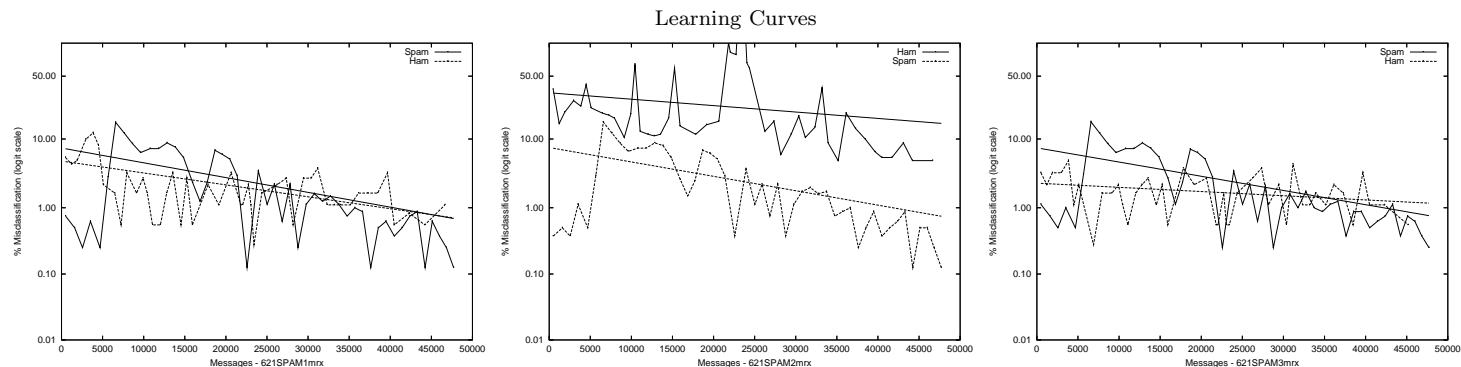
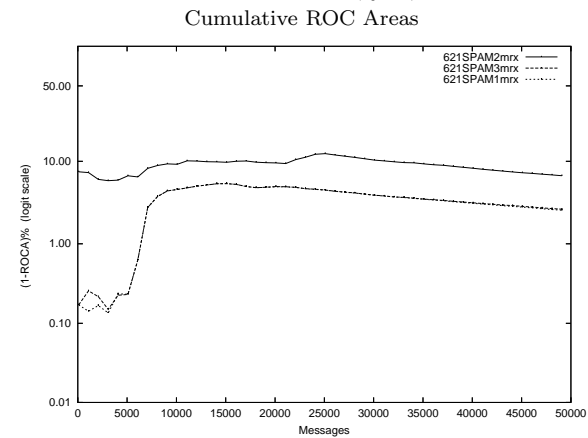
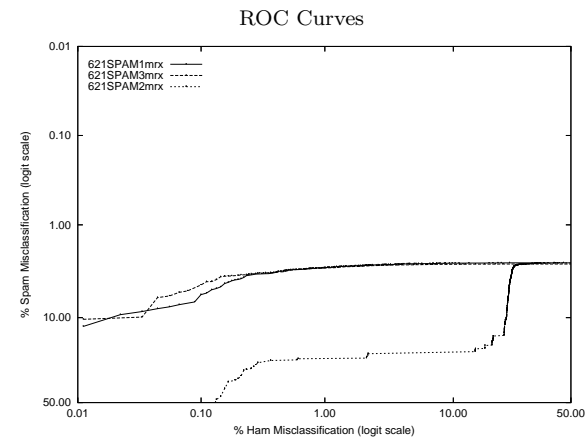
621 - IBM Research (Segal) on Public trec05p-1/full Corpus

Gold Std. Filter Classification	621SPAM1full		621SPAM2full		621SPAM3full	
	ham	spam	ham	spam	ham	spam
ham	38461	103	17705	564	38163	89
spam	938	52687	21694	52226	1236	52701
Ham Misc%	2.38 (2.23-2.54)		55.06 (54.57-55.55)		3.14 (2.97-3.31)	
Spam Misc%	0.20 (0.16-0.24)		1.07 (0.98-1.16)		0.17 (0.14-0.21)	
Logit av Misc%	0.69 (0.62-0.76)		10.32 (9.93-10.72)		0.73 (0.66-0.82)	
(1-ROCA)%	0.044 (0.034 - 0.058)		11.362 (11.152 - 11.576)		0.060 (0.048 - 0.074)	
Ham Misc%	Spam Misc%					
0.01	15.10 (6.89 - 29.94)		28.87 (21.46 - 37.63)		24.73 (14.26 - 39.37)	
0.02	11.99 (8.78 - 16.16)		28.87 (28.41 - 29.35)		17.69 (13.73 - 22.49)	
0.05	6.00 (4.27 - 8.38)		28.86 (28.39 - 29.33)		12.09 (8.83 - 16.34)	
0.10	3.63 (2.85 - 4.62)		28.85 (28.39 - 29.32)		7.02 (5.62 - 8.74)	
0.20	2.02 (1.73 - 2.35)		28.83 (28.36 - 29.30)		4.10 (3.51 - 4.79)	
0.50	0.94 (0.78 - 1.13)		28.75 (28.28 - 29.23)		1.62 (1.38 - 1.90)	
1.00	0.47 (0.41 - 0.55)		28.60 (28.13 - 29.09)		0.72 (0.60 - 0.86)	
2.00	0.24 (0.19 - 0.29)		28.25 (27.78 - 28.72)		0.29 (0.24 - 0.34)	
5.00	0.09 (0.07 - 0.13)		27.32 (26.86 - 27.79)		0.10 (0.08 - 0.14)	
10.00	0.04 (0.03 - 0.07)		26.14 (25.68 - 26.60)		0.04 (0.02 - 0.06)	
20.00	0.02 (0.01 - 0.04)		24.92 (24.49 - 25.36)		0.02 (0.01 - 0.04)	
50.00	0.01 (0.01 - 0.03)		13.32 (13.00 - 13.65)		0.02 (0.01 - 0.03)	
Spam Misc%	Ham Misc%					
0.01	99.48 (44.18 - 100.00)		99.48 (98.59 - 99.81)		99.48 (45.30 - 100.00)	
0.02	24.78 (0.63 - 94.46)		98.77 (98.10 - 99.21)		22.88 (0.54 - 94.16)	
0.05	8.98 (6.36 - 12.54)		97.93 (97.17 - 98.49)		8.22 (6.23 - 10.77)	
0.10	4.62 (3.58 - 5.93)		96.84 (96.06 - 97.47)		5.22 (4.17 - 6.53)	
0.20	2.31 (1.92 - 2.77)		92.63 (91.26 - 93.81)		2.80 (2.29 - 3.41)	
0.50	0.92 (0.76 - 1.12)		81.93 (79.74 - 83.93)		1.25 (1.09 - 1.42)	
1.00	0.47 (0.39 - 0.57)		55.26 (54.79 - 55.73)		0.72 (0.62 - 0.84)	
2.00	0.20 (0.15 - 0.26)		53.80 (53.32 - 54.27)		0.41 (0.35 - 0.48)	
5.00	0.07 (0.05 - 0.10)		52.58 (52.10 - 53.06)		0.14 (0.11 - 0.19)	
10.00	0.03 (0.01 - 0.05)		52.09 (51.62 - 52.56)		0.06 (0.04 - 0.09)	
20.00	0.01 (0.00 - 0.02)		28.10 (27.37 - 28.84)		0.01 (0.00 - 0.04)	
50.00	0.00 (0.00 - 0.01)		0.00 (0.00 - 0.01)		0.00 (0.00 - 0.01)	



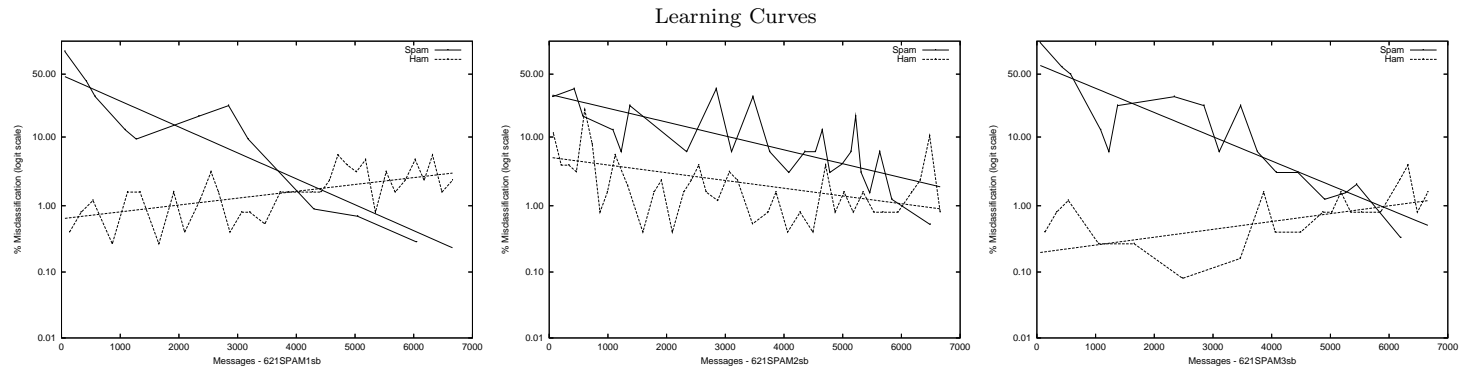
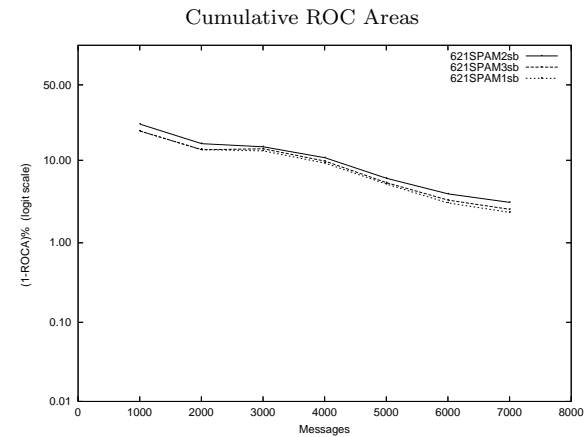
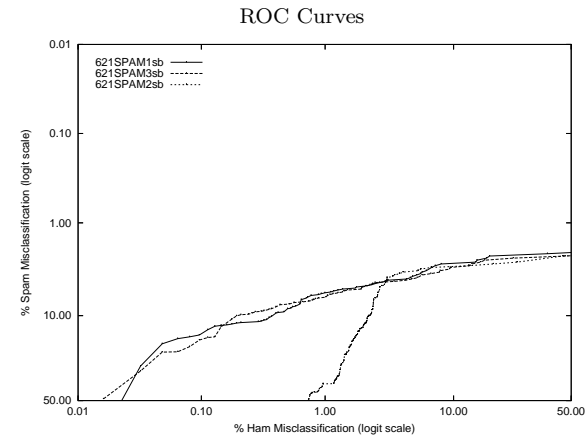
621 - IBM Research (Segal) on Private Mr X. Corpus

Gold Std.	621SPAM1mrx		621SPAM2mrx		621SPAM3mrx	
	ham	spam	ham	spam	ham	spam
Filter Classification	ham	8829 1109	ham	6704 1151	ham	8882 1151
	spam	209 38939	spam	2334 38897	spam	156 38897
Ham Misc%	2.31 (2.01-2.64)		25.82 (24.92-26.74)		1.73 (1.47-2.02)	
Spam Misc%	2.77 (2.61-2.93)		2.87 (2.71-3.04)		2.87 (2.71-3.04)	
Logit av Misc%	2.53 (2.35-2.72)		9.21 (8.90-9.53)		2.23 (2.05-2.42)	
(1-ROCA)%	2.616 (2.470 - 2.770)		6.814 (6.567 - 7.070)		2.692 (2.556 - 2.835)	
Ham Misc%	Spam Misc%					
0.01	14.16 (10.44 - 18.94)		94.19 (74.46 - 98.90)		17.01 (10.37 - 26.64)	
0.02	12.13 (7.86 - 18.25)		87.04 (64.78 - 96.08)		10.36 (6.04 - 17.18)	
0.05	8.05 (5.73 - 11.19)		76.30 (56.87 - 88.72)		6.21 (3.34 - 11.26)	
0.10	5.71 (3.72 - 8.67)		59.16 (32.87 - 81.08)		4.55 (3.30 - 6.23)	
0.20	3.96 (3.02 - 5.18)		34.44 (23.64 - 47.13)		3.59 (3.21 - 4.01)	
0.50	3.14 (2.90 - 3.40)		25.08 (24.43 - 25.74)		3.20 (3.00 - 3.41)	
1.00	2.96 (2.80 - 3.14)		24.24 (23.80 - 24.69)		2.99 (2.85 - 3.15)	
2.00	2.81 (2.66 - 2.97)		24.08 (22.65 - 25.57)		2.86 (2.72 - 3.00)	
5.00	2.69 (2.54 - 2.84)		22.01 (21.58 - 22.45)		2.76 (2.62 - 2.90)	
10.00	2.65 (2.50 - 2.80)		21.93 (21.49 - 22.37)		2.72 (2.59 - 2.87)	
20.00	2.63 (2.48 - 2.79)		15.13 (14.78 - 15.50)		2.71 (2.57 - 2.85)	
50.00	2.62 (2.48 - 2.78)		2.63 (2.49 - 2.78)		2.70 (2.57 - 2.85)	
Spam Misc%	Ham Misc%					
0.01	99.99 (nan - nan)		99.99 (nan - nan)		99.99 (nan - nan)	
0.02	99.96 (nan - nan)		99.96 (nan - nan)		99.97 (nan - nan)	
0.05	99.92 (nan - nan)		99.92 (nan - nan)		99.91 (99.74 - 99.97)	
0.10	99.88 (99.76 - 99.94)		99.88 (99.75 - 99.94)		99.87 (99.72 - 99.94)	
0.20	99.77 (99.59 - 99.87)		99.77 (99.58 - 99.87)		99.77 (99.59 - 99.87)	
0.50	99.47 (99.22 - 99.64)		99.47 (99.25 - 99.62)		99.48 (99.26 - 99.64)	
1.00	99.14 (98.88 - 99.34)		99.13 (98.87 - 99.32)		99.16 (98.89 - 99.36)	
2.00	98.43 (98.14 - 98.68)		98.43 (98.13 - 98.68)		98.51 (98.20 - 98.76)	
5.00	0.13 (0.07 - 0.26)		23.94 (23.10 - 24.81)		0.10 (0.05 - 0.22)	
10.00	0.02 (0.01 - 0.09)		23.08 (22.27 - 23.91)		0.03 (0.01 - 0.12)	
20.00	0.01 (0.01 - 0.01)		14.39 (12.73 - 16.24)		0.01 (0.01 - 0.01)	
50.00	0.01 (0.01 - 0.01)		0.13 (0.07 - 0.25)		0.01 (0.01 - 0.01)	



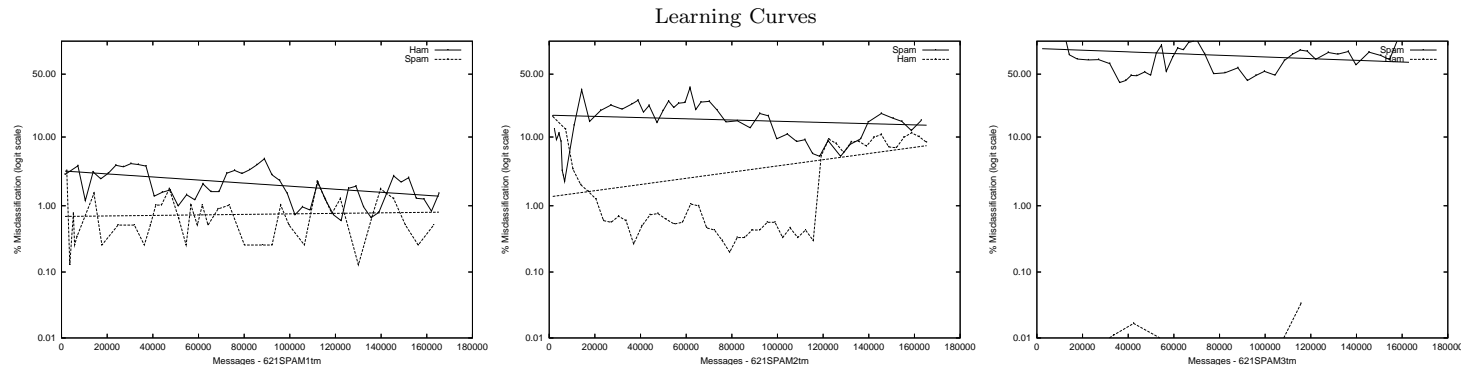
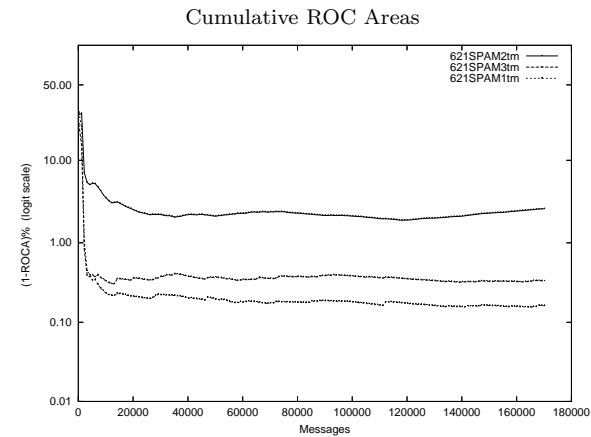
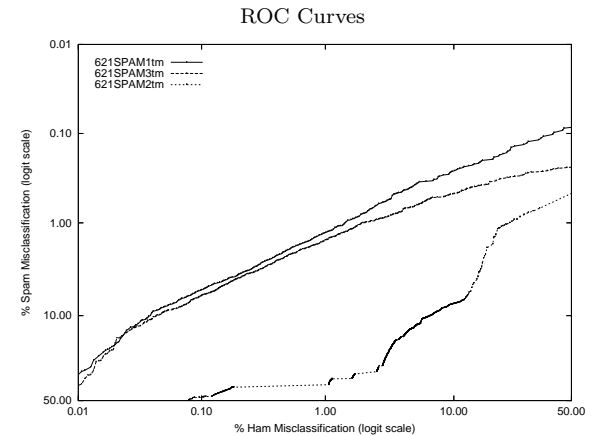
621 - IBM Research (Segal) on Private S.B. Corpus

Gold Std. Filter Classification	621SPAM1sb		621SPAM2sb		621SPAM3sb	
	ham	spam	ham	spam	ham	spam
Ham Misc%	6133	41	6077	53	6196	58
Spam Misc%	98	734	154	722	35	717
Logit av Misc%	1.57 (1.28-1.91)		2.47 (2.10-2.89)		0.56 (0.39-0.78)	
	5.29 (3.82-7.11)		6.84 (5.16-8.85)		7.48 (5.73-9.57)	
	2.90 (2.42-3.47)		4.13 (3.54-4.82)		2.09 (1.70-2.58)	
(1-ROCA)%	2.389 (1.630 - 3.490)		3.169 (2.269 - 4.409)		2.604 (1.766 - 3.825)	
Ham Misc%	Spam Misc%					
0.01	70.84 (43.56 - 88.43)		83.23 (75.95 - 88.63)		80.00 (42.38 - 95.61)	
0.02	70.06 (25.31 - 94.17)		80.65 (69.70 - 88.30)		48.90 (11.66 - 87.40)	
0.05	18.45 (4.16 - 54.12)		74.06 (59.17 - 84.91)		22.19 (7.64 - 49.58)	
0.10	15.48 (10.57 - 22.12)		61.94 (47.78 - 74.32)		17.16 (10.73 - 26.32)	
0.20	11.87 (9.39 - 14.90)		55.48 (49.48 - 61.33)		9.94 (6.40 - 15.12)	
0.50	8.52 (6.28 - 11.46)		54.84 (51.03 - 58.59)		7.61 (5.88 - 9.81)	
1.00	5.81 (4.35 - 7.71)		39.35 (33.49 - 45.54)		6.45 (4.81 - 8.60)	
2.00	5.03 (3.73 - 6.76)		13.29 (9.48 - 18.32)		5.03 (3.58 - 7.03)	
5.00	3.87 (2.81 - 5.30)		3.48 (2.33 - 5.17)		4.13 (2.88 - 5.89)	
10.00	2.84 (1.96 - 4.10)		3.10 (2.03 - 4.71)		3.10 (2.08 - 4.59)	
20.00	2.32 (1.53 - 3.52)		2.84 (1.85 - 4.33)		2.58 (1.70 - 3.91)	
50.00	2.19 (1.42 - 3.38)		2.32 (1.42 - 3.77)		2.32 (1.50 - 3.58)	
Spam Misc%	Ham Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.05	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.10	100.00 (nan - nan)		100.00 (nan - nan)		99.98 (nan - nan)	
0.20	100.00 (nan - nan)		100.00 (nan - nan)		99.98 (nan - nan)	
0.50	99.97 (99.89 - 99.99)		99.97 (99.90 - 99.99)		99.95 (99.84 - 99.99)	
1.00	99.47 (98.83 - 99.76)		99.47 (97.76 - 99.88)		99.25 (98.32 - 99.66)	
2.00	64.39 (0.72 - 99.78)		71.35 (1.69 - 99.72)		98.89 (39.10 - 99.99)	
5.00	2.09 (0.72 - 5.93)		2.82 (1.97 - 4.04)		2.15 (0.69 - 6.46)	
10.00	0.40 (0.21 - 0.78)		2.29 (1.94 - 2.72)		0.19 (0.09 - 0.41)	
20.00	0.05 (0.02 - 0.14)		1.57 (1.31 - 1.89)		0.08 (0.03 - 0.22)	
50.00	0.03 (0.01 - 0.11)		0.74 (0.58 - 0.94)		0.02 (0.00 - 0.06)	



621 - IBM Research (Segal) on Private T.M. Corpus

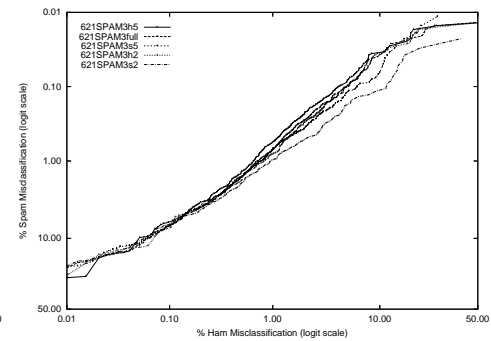
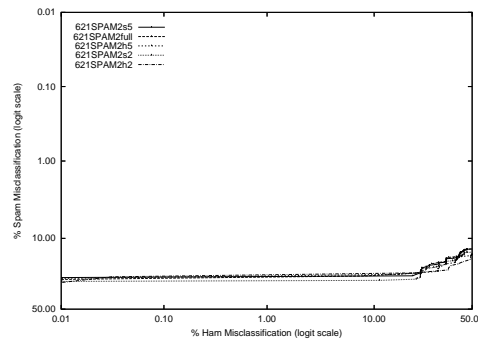
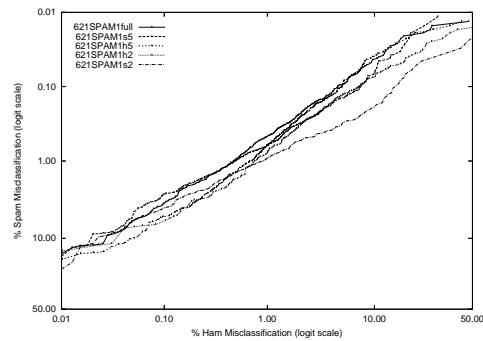
Gold Std. Filter Classification	621SPAM1tm			621SPAM2tm			621SPAM3tm		
	ham	spam	spam	ham	spam	spam	ham	spam	
Filter	ham	147412	144	ham	144916	3322	ham	150678	12933
Classification	spam	3273	19372	spam	5769	16194	spam	7	6583
Ham Misc%	2.17 (2.10-2.25)			3.83 (3.73-3.93)			0.00 (0.00-0.01)		
Spam Misc%	0.74 (0.62-0.87)			17.02 (16.50-17.56)			66.27 (65.60-66.93)		
Logit av Misc%	1.27 (1.17-1.38)			8.29 (8.12-8.46)			0.95 (0.66-1.37)		
(1-ROCA)%	0.161 (0.130 - 0.200)			2.647 (2.554 - 2.742)			0.332 (0.269 - 0.409)		
Ham Misc%	Spam Misc%								
0.01	33.86 (17.40 - 55.44)			90.87 (83.63 - 95.10)			40.27 (25.31 - 57.28)		
0.02	18.74 (12.48 - 27.18)			71.23 (54.93 - 83.42)			18.35 (11.80 - 27.42)		
0.05	8.08 (6.95 - 9.36)			54.96 (51.88 - 58.00)			9.11 (8.07 - 10.27)		
0.10	5.42 (4.79 - 6.12)			47.89 (46.93 - 48.86)			6.15 (5.60 - 6.76)		
0.20	3.67 (3.31 - 4.08)			41.38 (40.69 - 42.06)			4.05 (3.65 - 4.49)		
0.50	1.99 (1.78 - 2.22)			40.83 (40.21 - 41.46)			2.29 (2.07 - 2.54)		
1.00	1.28 (1.13 - 1.45)			39.98 (39.36 - 40.61)			1.55 (1.37 - 1.75)		
2.00	0.79 (0.68 - 0.92)			33.01 (32.36 - 33.67)			1.00 (0.87 - 1.15)		
5.00	0.38 (0.31 - 0.48)			13.25 (12.67 - 13.85)			0.66 (0.55 - 0.78)		
10.00	0.26 (0.20 - 0.34)			7.56 (7.26 - 7.86)			0.48 (0.39 - 0.58)		
20.00	0.18 (0.14 - 0.24)			1.21 (1.07 - 1.38)			0.34 (0.26 - 0.43)		
50.00	0.09 (0.05 - 0.15)			0.49 (0.40 - 0.60)			0.24 (0.18 - 0.32)		
Spam Misc%	Ham Misc%								
0.01	99.99 (nan - nan)			99.99 (nan - nan)			100.00 (nan - nan)		
0.02	99.98 (91.82 - 100.00)			99.72 (82.44 - 100.00)			99.99 (99.98 - 100.00)		
0.05	76.25 (44.38 - 92.82)			95.98 (91.47 - 98.16)			99.96 (99.91 - 99.98)		
0.10	41.84 (23.60 - 62.62)			89.66 (82.31 - 94.17)			99.93 (99.87 - 99.96)		
0.20	15.67 (9.86 - 23.99)			74.05 (65.56 - 81.04)			79.95 (2.37 - 99.85)		
0.50	3.38 (2.63 - 4.33)			49.45 (41.84 - 57.08)			9.03 (5.65 - 14.13)		
1.00	1.44 (1.19 - 1.73)			23.94 (20.73 - 27.48)			1.99 (1.47 - 2.68)		
2.00	0.50 (0.43 - 0.57)			16.95 (16.49 - 17.41)			0.62 (0.52 - 0.75)		
5.00	0.11 (0.09 - 0.14)			13.51 (13.30 - 13.74)			0.15 (0.12 - 0.17)		
10.00	0.04 (0.03 - 0.05)			6.75 (6.35 - 7.18)			0.04 (0.03 - 0.05)		
20.00	0.02 (0.01 - 0.03)			3.35 (3.26 - 3.45)			0.02 (0.01 - 0.03)		
50.00	0.01 (0.00 - 0.02)			0.08 (0.06 - 0.10)			0.01 (0.00 - 0.01)		



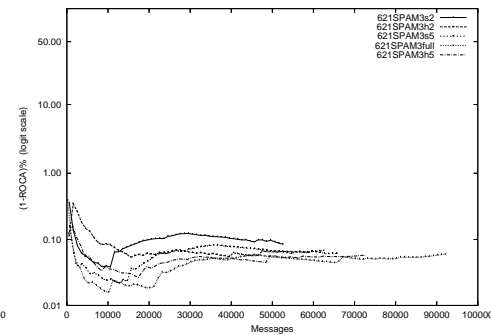
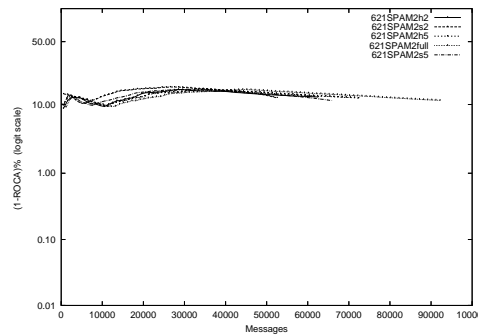
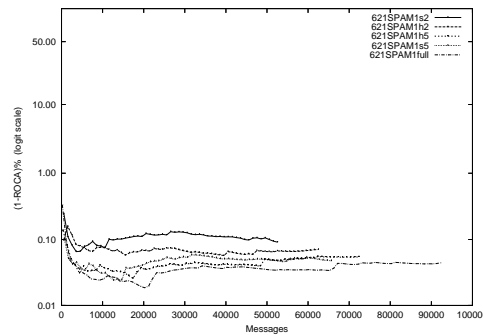
621 - IBM Research (Segal) - Public subset Corpora

	621SPAM1full	621SPAM1h2	621SPAM1h5	621SPAM1s2	621SPAM1s5
Ham Misc%	2.38 (2.23-2.54)	3.94 (3.56-4.34)	2.78 (2.56-3.02)	3.45 (3.27-3.64)	2.51 (2.36-2.67)
Spam Misc%	0.20 (0.16-0.24)	0.17 (0.14-0.21)	0.19 (0.15-0.23)	0.42 (0.32-0.55)	0.27 (0.21-0.34)
Logit av Misc%	0.69 (0.62-0.76)	0.83 (0.74-0.93)	0.72 (0.65-0.81)	1.22 (1.07-1.39)	0.83 (0.74-0.94)
(1-ROCA)%	0.044 (0.034 - 0.058)	0.070 (0.050 - 0.100)	0.054 (0.045 - 0.066)	0.091 (0.066 - 0.125)	0.048 (0.034 - 0.066)
	621SPAM2full	621SPAM2h2	621SPAM2h5	621SPAM2s2	621SPAM2s5
Ham Misc%	55.06 (54.57-55.55)	58.82 (57.84-59.80)	57.10 (56.40-57.79)	54.53 (54.03-55.02)	54.80 (54.30-55.29)
Spam Misc%	1.07 (0.98-1.16)	1.39 (1.29-1.49)	1.18 (1.09-1.28)	1.34 (1.15-1.55)	0.86 (0.76-0.98)
Logit av Misc%	10.32 (9.93-10.72)	12.42 (11.98-12.88)	11.20 (10.79-11.63)	11.33 (10.60-12.10)	9.32 (8.78-9.89)
(1-ROCA)%	11.362 (11.152 - 11.576)	12.626 (12.329 - 12.929)	12.221 (11.992 - 12.455)	12.291 (11.957 - 12.634)	11.352 (11.128 - 11.579)
	621SPAM3full	621SPAM3h2	621SPAM3h5	621SPAM3s2	621SPAM3s5
Ham Misc%	3.14 (2.97-3.31)	3.33 (2.99-3.71)	2.84 (2.62-3.09)	5.28 (5.07-5.51)	4.32 (4.12-4.52)
Spam Misc%	0.17 (0.14-0.21)	0.16 (0.13-0.20)	0.16 (0.13-0.20)	0.17 (0.11-0.26)	0.16 (0.12-0.22)
Logit av Misc%	0.73 (0.66-0.82)	0.74 (0.66-0.84)	0.68 (0.61-0.76)	0.98 (0.80-1.20)	0.84 (0.72-0.98)
(1-ROCA)%	0.060 (0.048 - 0.074)	0.068 (0.049 - 0.094)	0.058 (0.043 - 0.076)	0.085 (0.059 - 0.122)	0.061 (0.048 - 0.079)

ROC Curves

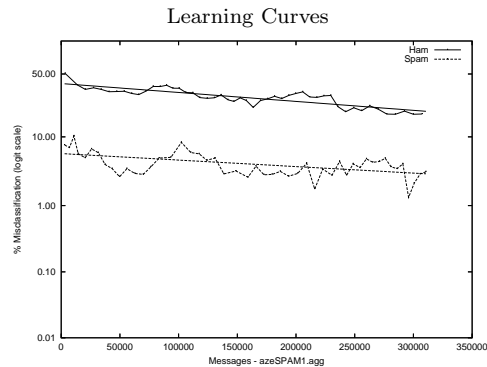
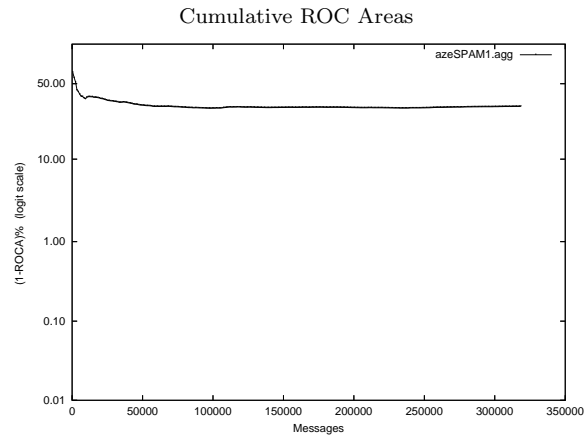
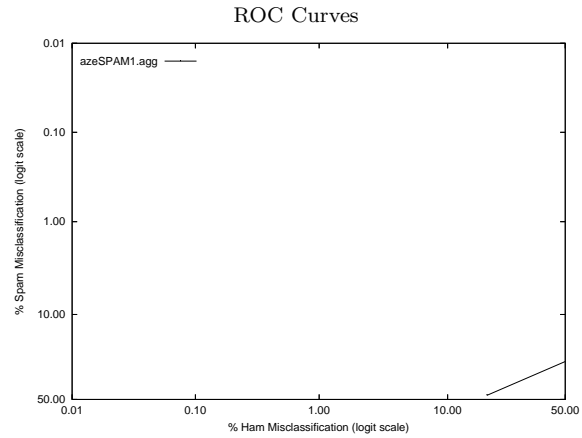


Cumulative ROC Area



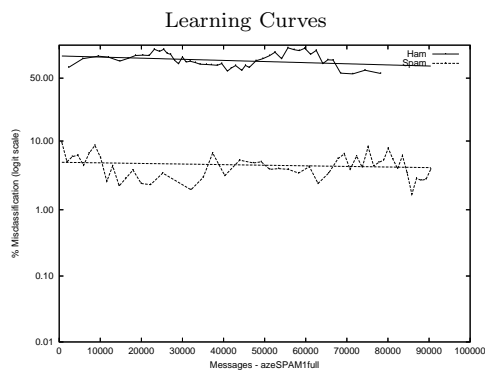
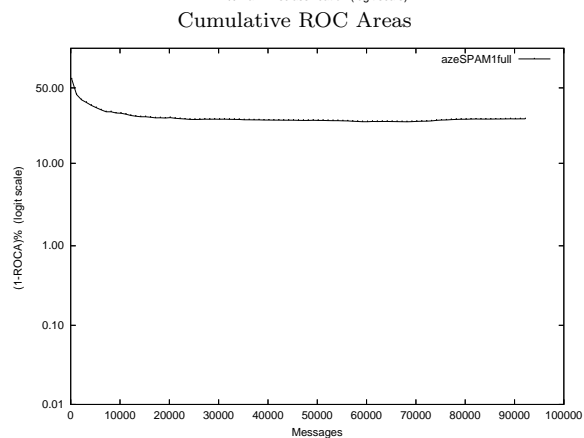
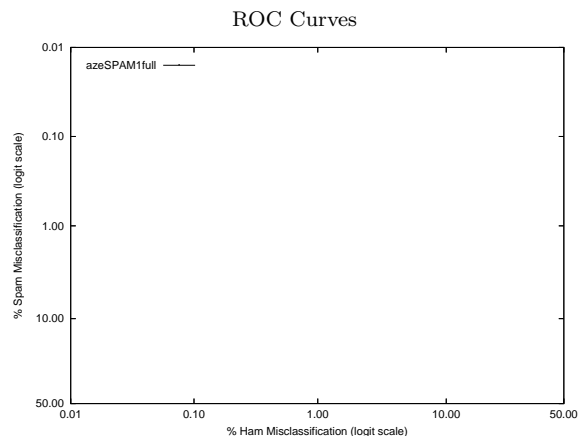
aze - U. Paris - Sud - Aggregate of full, mrx, sb, tm Corpus

Gold Std.	azeSPAM1.agg	
	ham	spam
Filter	142139	4760
Classification	spam	63214 108369
Ham Misc%	30.78 (30.58-30.98)	
Spam Misc%	4.21 (4.09-4.33)	
Logit av Misc%	12.26 (12.10-12.43)	
(1-ROCA)%	34.079 (33.876 - 34.282)	
Ham Misc%	Spam Misc%	
0.01	99.98 (99.97 - 99.99)	
0.02	99.96 (99.94 - 99.97)	
0.05	99.89 (99.86 - 99.92)	
0.10	99.76 (99.72 - 99.80)	
0.20	99.49 (99.43 - 99.55)	
0.50	98.60 (98.49 - 98.71)	
1.00	97.29 (97.15 - 97.43)	
2.00	94.57 (94.35 - 94.78)	
5.00	86.02 (85.73 - 86.31)	
10.00	72.11 (71.72 - 72.50)	
20.00	46.64 (46.34 - 46.95)	
50.00	30.52 (30.23 - 30.82)	
Spam Misc%	Ham Misc%	
0.01	99.99 (99.98 - 99.99)	
0.02	99.97 (99.95 - 99.98)	
0.05	99.91 (99.88 - 99.94)	
0.10	99.84 (99.80 - 99.87)	
0.20	99.66 (99.61 - 99.71)	
0.50	99.19 (99.10 - 99.27)	
1.00	98.50 (98.39 - 98.60)	
2.00	97.16 (97.01 - 97.30)	
5.00	92.71 (92.49 - 92.93)	
10.00	85.70 (85.41 - 85.98)	
20.00	69.12 (68.68 - 69.56)	
50.00	17.91 (17.73 - 18.10)	



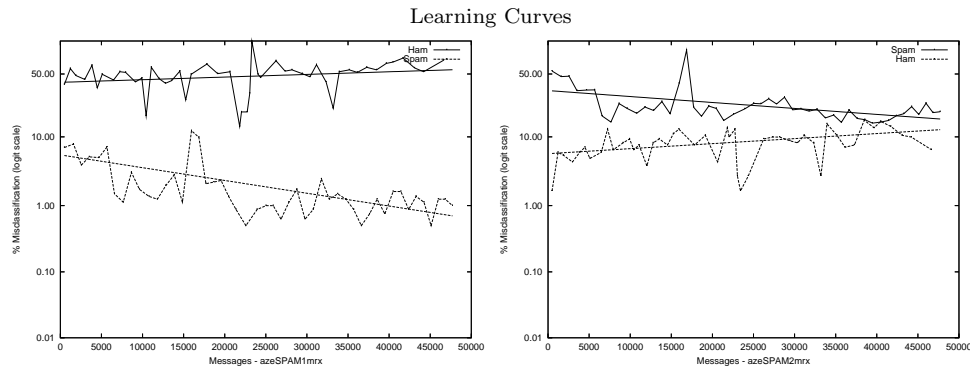
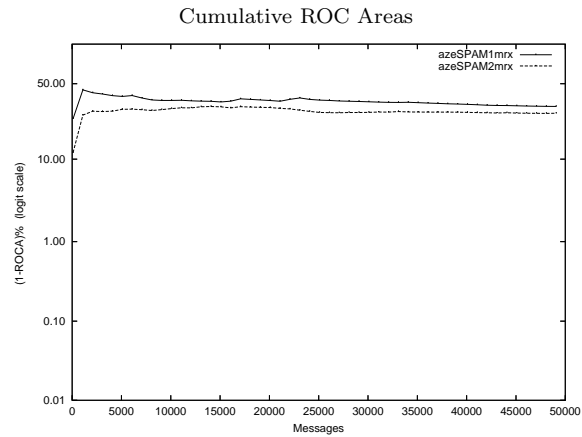
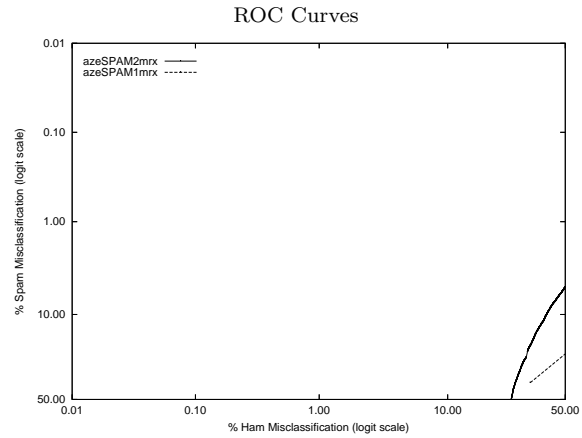
aze - U. Paris - Sud - Public trec05p-1/full Corpus

Gold Std.	azeSPAM1full	
	ham	spam
Filter	ham	13851 2414
Classification	spam	25548 50376
Ham Misc%	64.84 (64.37-65.32)	
Spam Misc%	4.57 (4.40-4.75)	
Logit av Misc%	22.92 (22.52-23.32)	
(1-ROCA)%	28.887 (28.489 - 29.288)	
Ham Misc%	Spam Misc%	
0.01	99.98 (99.94 - 99.99)	
0.02	99.94 (99.88 - 99.97)	
0.05	99.81 (99.68 - 99.89)	
0.10	99.50 (99.29 - 99.65)	
0.20	98.87 (98.54 - 99.12)	
0.50	97.50 (97.10 - 97.85)	
1.00	94.44 (93.82 - 95.00)	
2.00	88.82 (88.00 - 89.59)	
5.00	70.27 (68.83 - 71.68)	
10.00	49.65 (49.10 - 50.20)	
20.00	43.90 (43.36 - 44.44)	
50.00	26.84 (26.33 - 27.36)	
Spam Misc%	Ham Misc%	
0.01	99.98 (99.94 - 99.99)	
0.02	99.97 (99.93 - 99.99)	
0.05	99.88 (99.82 - 99.92)	
0.10	99.81 (99.75 - 99.86)	
0.20	99.62 (99.53 - 99.70)	
0.50	99.01 (98.87 - 99.13)	
1.00	98.11 (97.92 - 98.29)	
2.00	96.09 (95.77 - 96.39)	
5.00	90.13 (89.71 - 90.54)	
10.00	80.46 (79.85 - 81.06)	
20.00	61.76 (60.88 - 62.62)	
50.00	9.33 (8.42 - 10.32)	



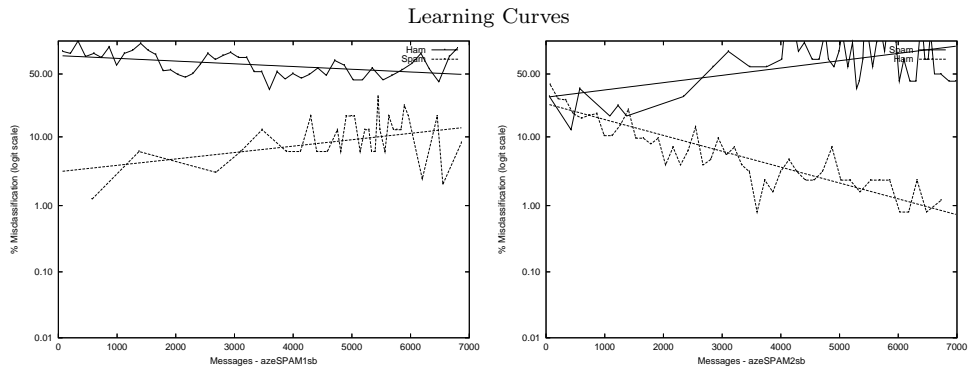
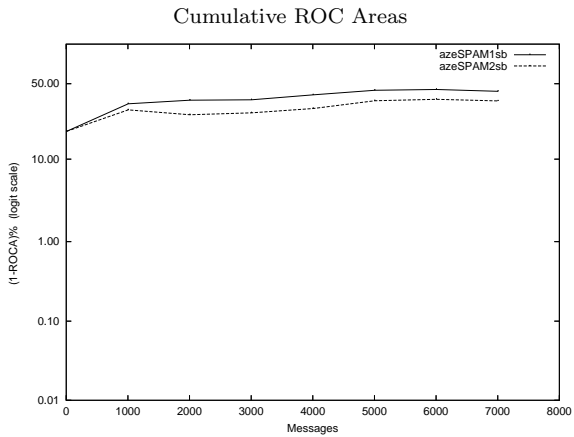
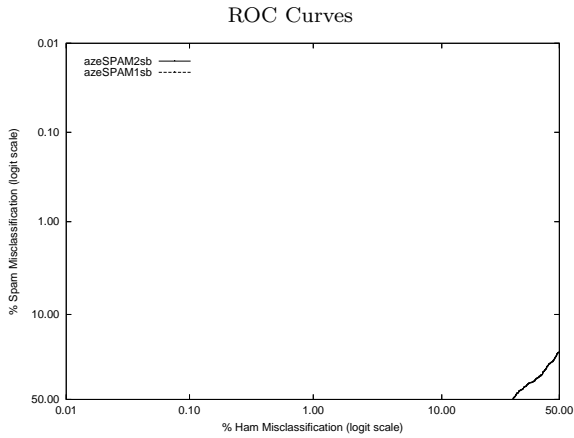
aze - U. Paris - Sud - Private Mr X. Corpus

Gold Std. Filter Classification	azeSPAM1mrX		azeSPAM2mrX	
	ham	spam	ham	spam
Ham Misc%	47.81 (46.77-48.85)		8.54 (7.97-9.14)	
Spam Misc%	2.28 (2.14-2.43)		25.35 (24.93-25.78)	
Logit av Misc%	12.76 (12.34-13.20)		15.12 (14.63-15.62)	
(1-ROCA)%	34.048 (33.364 - 34.739)		29.765 (28.951 - 30.592)	
Ham Misc%	Spam Misc%			
0.01	100.00 (nan - nan)		100.00 (nan - nan)	
0.02	99.98 (nan - nan)		99.99 (nan - nan)	
0.05	99.84 (99.63 - 99.93)		99.97 (nan - nan)	
0.10	99.69 (99.44 - 99.83)		99.95 (99.86 - 99.98)	
0.20	99.53 (99.27 - 99.70)		99.93 (99.84 - 99.97)	
0.50	98.96 (98.59 - 99.23)		99.80 (99.71 - 99.86)	
1.00	97.92 (97.44 - 98.32)		99.65 (99.54 - 99.73)	
2.00	95.99 (95.43 - 96.49)		99.33 (99.17 - 99.46)	
5.00	90.30 (89.44 - 91.09)		98.52 (98.31 - 98.70)	
10.00	81.93 (80.65 - 83.14)		97.10 (96.85 - 97.33)	
20.00	64.06 (62.35 - 65.74)		94.15 (93.82 - 94.47)	
50.00	26.59 (25.67 - 27.52)		5.20 (4.73 - 5.71)	
Spam Misc%	Ham Misc%			
0.01	99.87 (nan - nan)		99.68 (99.02 - 99.90)	
0.02	99.86 (nan - nan)		99.37 (98.60 - 99.72)	
0.05	99.77 (99.47 - 99.90)		98.14 (97.25 - 98.75)	
0.10	99.68 (99.47 - 99.81)		96.20 (94.86 - 97.21)	
0.20	99.48 (99.26 - 99.64)		92.92 (91.27 - 94.28)	
0.50	98.46 (98.05 - 98.79)		84.48 (82.13 - 86.57)	
1.00	96.91 (96.31 - 97.42)		72.98 (71.64 - 74.28)	
2.00	94.28 (93.64 - 94.86)		63.50 (62.27 - 64.71)	
5.00	85.31 (84.27 - 86.29)		50.59 (49.27 - 51.91)	
10.00	71.10 (69.92 - 72.25)		41.43 (40.37 - 42.49)	
20.00	58.38 (57.28 - 59.46)		34.20 (33.22 - 35.19)	
50.00	27.45 (26.44 - 28.49)		26.75 (25.86 - 27.67)	



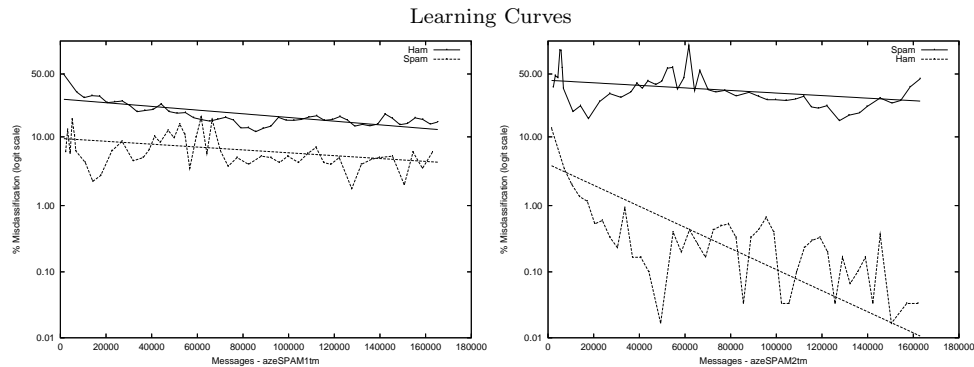
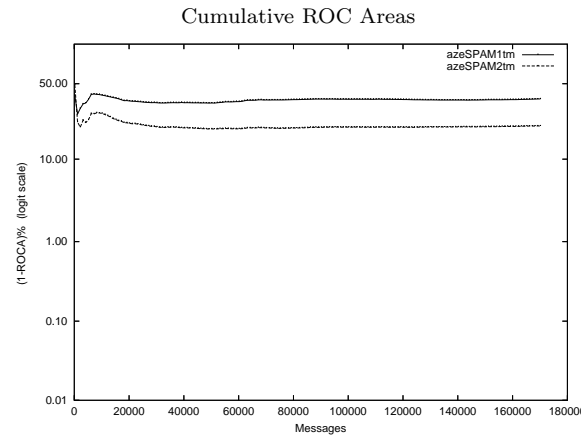
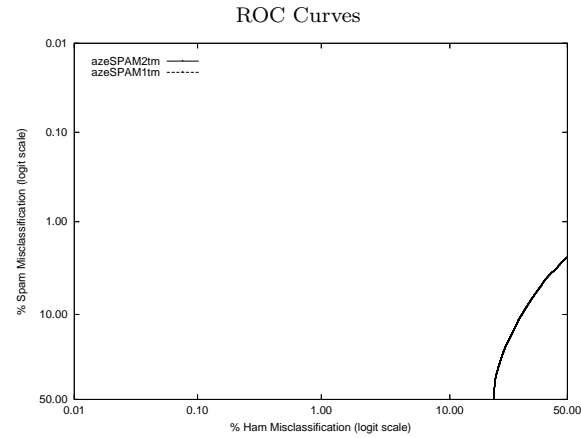
aze - U. Paris - Sud - Private S.B. Corpus

Gold Std.	azeSPAM1sb		azeSPAM2sb			
	ham	spam	ham	spam		
Filter	ham	2623	71	ham	5730	461
Classification	spam	3608	704	spam	501	314
Ham Misc%	57.90 (56.67-59.13)		8.04 (7.38-8.74)			
Spam Misc%	9.16 (7.22-11.42)		59.48 (55.93-62.96)			
Logit av Misc%	27.14 (24.75-29.67)		26.38 (24.76-28.06)			
(1-ROCA)%	44.502 (42.314 - 46.712)		37.739 (35.888 - 39.626)			
Ham Misc%	Spam Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)			
0.02	100.00 (nan - nan)		100.00 (nan - nan)			
0.05	99.87 (nan - nan)		100.00 (nan - nan)			
0.10	99.48 (nan - nan)		100.00 (nan - nan)			
0.20	99.35 (nan - nan)		100.00 (nan - nan)			
0.50	98.58 (95.96 - 99.51)		100.00 (nan - nan)			
1.00	97.94 (95.68 - 99.02)		100.00 (nan - nan)			
2.00	97.55 (96.01 - 98.50)		99.87 (nan - nan)			
5.00	93.68 (91.81 - 95.14)		99.87 (nan - nan)			
10.00	87.35 (84.40 - 89.82)		99.61 (nan - nan)			
20.00	73.16 (69.34 - 76.66)		72.90 (65.12 - 79.49)			
50.00	42.71 (39.05 - 46.46)		22.84 (19.78 - 26.22)			
Spam Misc%	Ham Misc%					
0.01	-100.00 (nan - nan)		-100.00 (nan - nan)			
0.02	-100.00 (nan - nan)		-100.00 (nan - nan)			
0.05	-100.00 (nan - nan)		-100.00 (nan - nan)			
0.10	99.97 (nan - nan)		99.89 (nan - nan)			
0.20	99.97 (nan - nan)		99.79 (nan - nan)			
0.50	99.78 (99.12 - 99.94)		99.71 (98.84 - 99.93)			
1.00	99.50 (98.82 - 99.79)		98.92 (97.56 - 99.53)			
2.00	98.57 (97.41 - 99.22)		97.05 (94.95 - 98.29)			
5.00	94.86 (92.76 - 96.38)		88.38 (83.50 - 91.96)			
10.00	87.96 (84.34 - 90.84)		74.39 (70.90 - 77.59)			
20.00	73.25 (68.89 - 77.20)		53.43 (49.22 - 57.58)			
50.00	42.66 (38.80 - 46.61)		29.66 (27.41 - 32.01)			



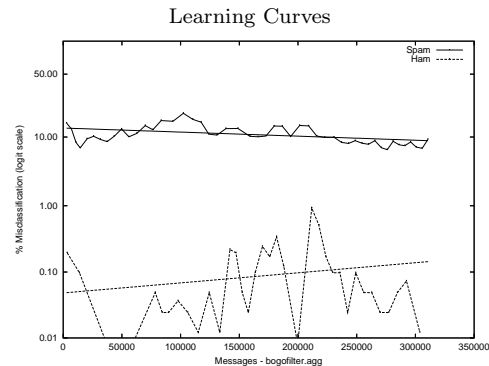
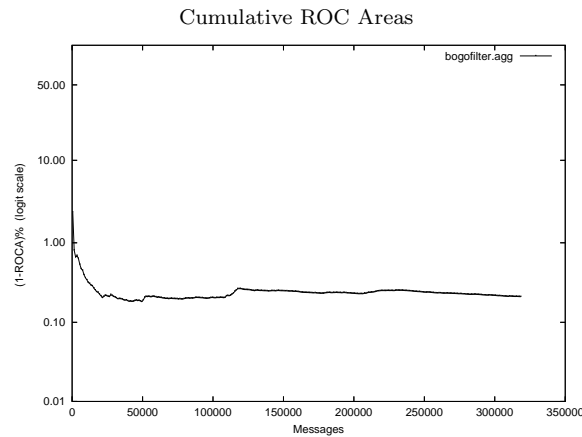
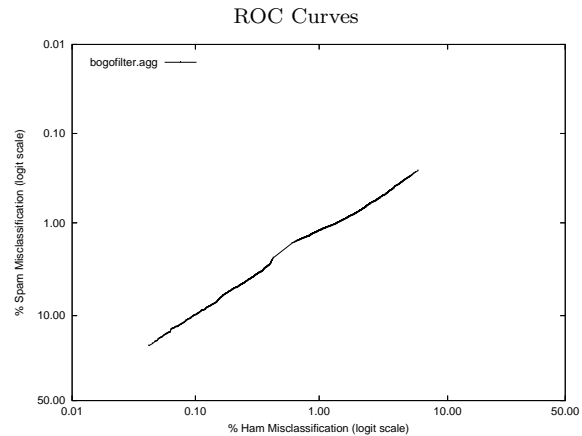
aze - U. Paris - Sud - Private T.M. Corpus

Gold Std. Filter Classification	azeSPAM1tm		azeSPAM2tm	
	ham spam	ham spam	ham spam	ham spam
Ham Misc%	19.73 (19.53-19.94)		0.63 (0.59-0.68)	
Spam Misc%	6.97 (6.62-7.34)		36.84 (36.16-37.52)	
Logit av Misc%	11.95 (11.66-12.25)		5.75 (5.56-5.94)	
(1-ROCA)%	39.082 (38.627 - 39.540)		22.625 (22.354 - 22.899)	
Ham Misc%	Spam Misc%			
0.01	99.98 (nan - nan)		99.99 (nan - nan)	
0.02	99.96 (99.92 - 99.98)		99.99 (nan - nan)	
0.05	99.86 (99.78 - 99.91)		99.93 (99.85 - 99.96)	
0.10	99.72 (99.62 - 99.80)		99.89 (99.83 - 99.92)	
0.20	99.49 (99.35 - 99.60)		99.76 (99.69 - 99.82)	
0.50	98.82 (98.62 - 99.00)		99.51 (99.40 - 99.60)	
1.00	97.75 (97.49 - 97.98)		99.06 (98.91 - 99.20)	
2.00	95.56 (95.20 - 95.90)		98.26 (98.05 - 98.46)	
5.00	88.67 (88.15 - 89.17)		95.94 (95.64 - 96.21)	
10.00	77.75 (77.07 - 78.41)		91.70 (91.30 - 92.08)	
20.00	54.88 (54.09 - 55.66)		82.92 (64.76 - 92.77)	
50.00	36.32 (35.65 - 37.00)		4.26 (3.95 - 4.58)	
Spam Misc%	Ham Misc%			
0.01	99.99 (nan - nan)		99.97 (99.82 - 99.99)	
0.02	99.98 (99.92 - 99.99)		99.89 (99.68 - 99.96)	
0.05	99.96 (99.90 - 99.98)		99.76 (99.57 - 99.86)	
0.10	99.93 (99.88 - 99.96)		99.13 (98.60 - 99.46)	
0.20	99.82 (99.74 - 99.88)		98.09 (97.34 - 98.62)	
0.50	99.49 (99.39 - 99.58)		94.06 (92.70 - 95.18)	
1.00	98.92 (98.76 - 99.06)		87.71 (86.02 - 89.22)	
2.00	97.89 (97.66 - 98.10)		76.16 (73.82 - 78.36)	
5.00	94.70 (94.36 - 95.02)		41.00 (37.30 - 44.81)	
10.00	89.37 (88.91 - 89.82)		30.08 (29.53 - 30.64)	
20.00	76.85 (75.94 - 77.74)		24.18 (23.87 - 24.50)	
50.00	26.69 (25.51 - 27.91)		20.21 (20.00 - 20.41)	



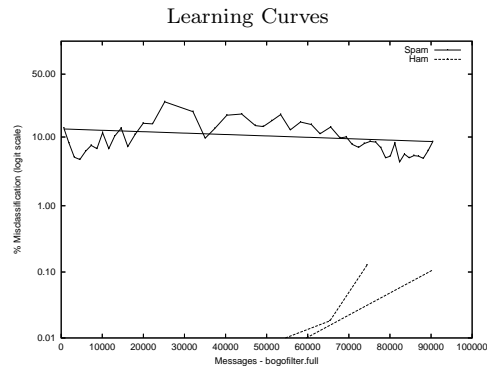
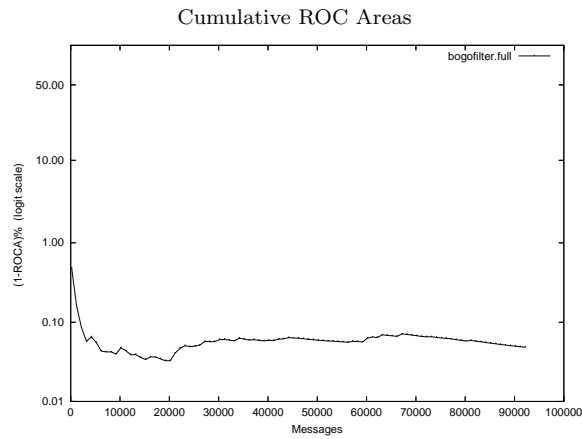
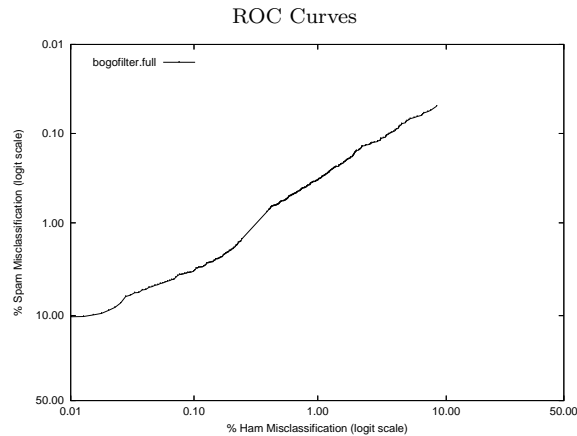
bogofilter - Bogofilter (non-participant) on Aggregate of full, mrx, sb, tm Corpus

Gold Std.	bogofilter.agg	
	ham	spam
Filter	ham	205173 12291
Classification	spam	180 100838
Ham Misc%	0.09 (0.08-0.10)	
Spam Misc%	10.86 (10.68-11.05)	
Logit av Misc%	1.02 (0.95-1.10)	
(1-ROCA)%	0.210 (0.193 - 0.228)	
Ham Misc%	Spam Misc%	
0.01	78.53 (70.67 - 84.74)	
0.02	59.50 (47.32 - 70.62)	
0.05	17.08 (13.97 - 20.72)	
0.10	9.86 (8.87 - 10.95)	
0.20	5.35 (4.97 - 5.76)	
0.50	2.10 (1.96 - 2.25)	
1.00	1.22 (1.15 - 1.29)	
2.00	0.77 (0.71 - 0.83)	
5.00	0.31 (0.29 - 0.35)	
10.00	0.25 (0.23 - 0.28)	
20.00	0.23 (0.21 - 0.26)	
50.00	0.14 (0.12 - 0.17)	
Spam Misc%	Ham Misc%	
0.01	95.80 (92.68 - 97.63)	
0.02	91.56 (87.54 - 94.36)	
0.05	83.11 (78.44 - 86.94)	
0.10	63.03 (56.45 - 69.16)	
0.20	33.36 (24.74 - 43.26)	
0.50	3.19 (2.90 - 3.50)	
1.00	1.39 (1.28 - 1.52)	
2.00	0.52 (0.49 - 0.56)	
5.00	0.22 (0.20 - 0.24)	
10.00	0.10 (0.08 - 0.11)	
20.00	0.04 (0.03 - 0.05)	
50.00	0.03 (0.02 - 0.03)	



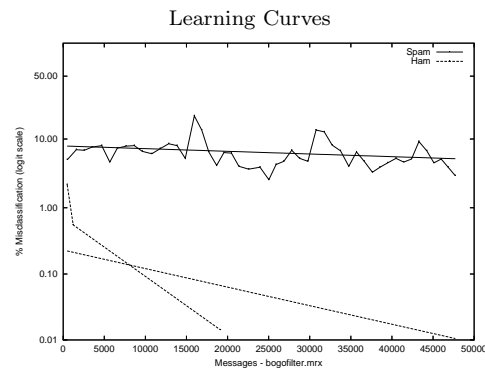
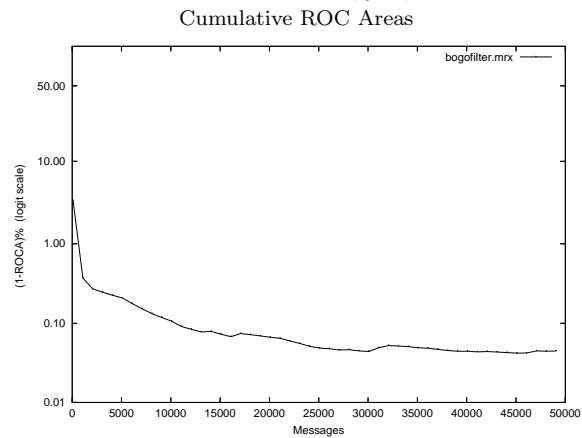
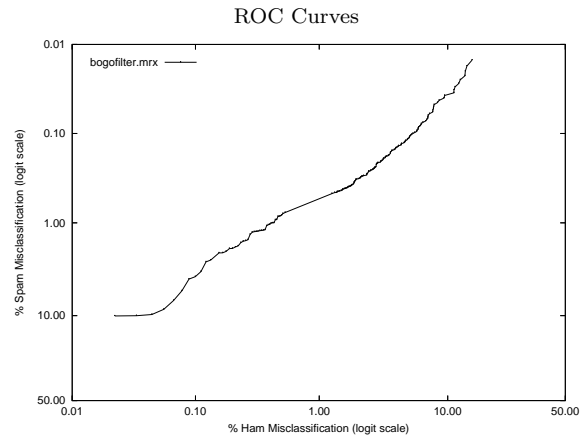
bogofilter - Bogofilter (non-participant) on Public trec05p-1/full Corpus

Gold Std.	bogofilter.full	
	ham	spam
Filter	39396	5525
Classification	spam	3 47265
Ham Misc%	0.01 (0.00-0.02)	
Spam Misc%	10.47 (10.21-10.73)	
Logit av Misc%	0.30 (0.17-0.52)	
(1-ROCA)%	0.048 (0.038 - 0.062)	
Ham Misc%	Spam Misc%	
0.01	10.39 (6.96 - 15.25)	
0.02	9.49 (6.70 - 13.26)	
0.05	4.83 (3.74 - 6.22)	
0.10	3.41 (2.83 - 4.10)	
0.20	2.02 (1.58 - 2.58)	
0.50	0.59 (0.50 - 0.68)	
1.00	0.33 (0.28 - 0.39)	
2.00	0.16 (0.13 - 0.20)	
5.00	0.07 (0.05 - 0.10)	
10.00	0.05 (0.03 - 0.07)	
20.00	0.05 (0.03 - 0.07)	
50.00	0.03 (0.02 - 0.05)	
Spam Misc%	Ham Misc%	
0.01	79.58 (58.02 - 91.66)	
0.02	72.47 (50.70 - 87.07)	
0.05	8.50 (1.71 - 33.18)	
0.10	3.78 (2.82 - 5.05)	
0.20	1.70 (1.39 - 2.07)	
0.50	0.61 (0.51 - 0.72)	
1.00	0.34 (0.28 - 0.40)	
2.00	0.20 (0.16 - 0.25)	
5.00	0.05 (0.03 - 0.08)	
10.00	0.02 (0.01 - 0.03)	
20.00	0.00 (0.00 - 0.01)	
50.00	0.00 (0.00 - 0.00)	



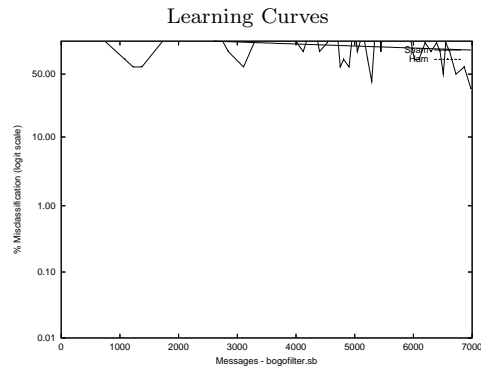
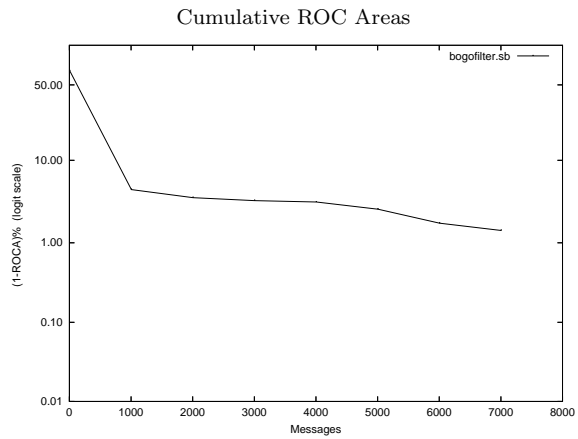
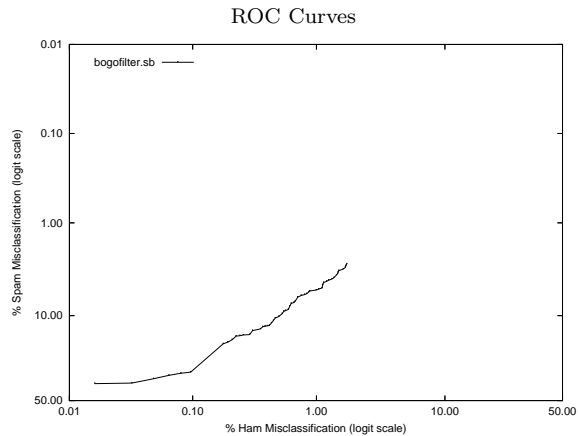
bogofilter - Bogofilter (non-participant) on Private Mr X. Corpus

Gold Std.	bogofilter.mrx	
	ham	spam
Filter	9031	2607
Classification	spam	7 37441
Ham Misc%	0.08 (0.03-0.16)	
Spam Misc%	6.51 (6.27-6.76)	
Logit av Misc%	0.73 (0.50-1.05)	
(1-ROCA)%	0.045 (0.032 - 0.063)	
Ham Misc%	Spam Misc%	
0.01	31.04 (6.26 - 75.21)	
0.02	25.16 (5.16 - 67.51)	
0.05	9.74 (3.85 - 22.52)	
0.10	3.90 (1.55 - 9.50)	
0.20	1.92 (1.25 - 2.93)	
0.50	0.81 (0.63 - 1.04)	
1.00	0.57 (0.47 - 0.70)	
2.00	0.32 (0.26 - 0.41)	
5.00	0.12 (0.09 - 0.16)	
10.00	0.04 (0.02 - 0.06)	
20.00	0.01 (0.01 - 0.03)	
50.00	0.01 (0.01 - 0.04)	
Spam Misc%	Ham Misc%	
0.01	56.52 (12.66 - 92.10)	
0.02	13.31 (4.06 - 35.79)	
0.05	7.88 (5.87 - 10.50)	
0.10	5.49 (4.35 - 6.90)	
0.20	3.21 (2.53 - 4.07)	
0.50	1.14 (0.87 - 1.49)	
1.00	0.43 (0.31 - 0.60)	
2.00	0.19 (0.11 - 0.33)	
5.00	0.09 (0.04 - 0.21)	
10.00	0.04 (0.01 - 0.20)	
20.00	0.02 (0.01 - 0.09)	
50.00	0.01 (0.00 - 0.02)	



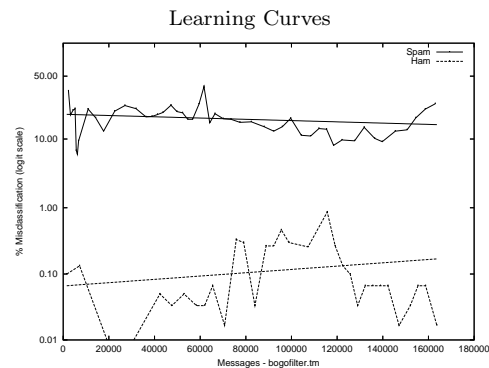
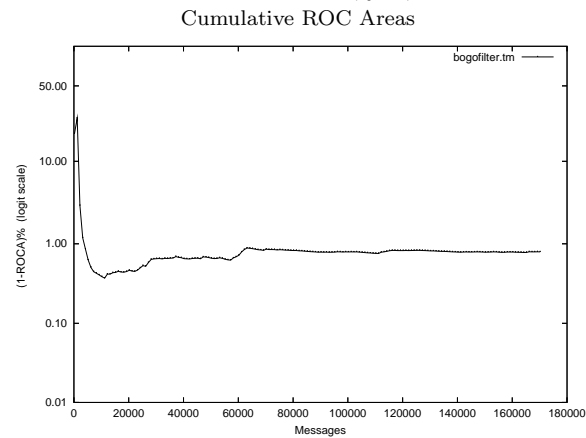
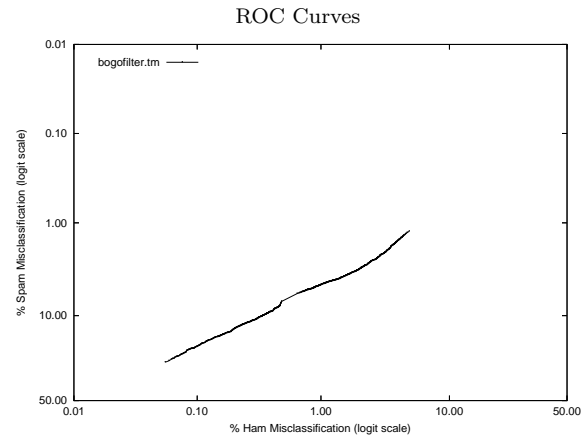
bogofilter - Bogofilter (non-participant) on Private S.B. Corpus

Gold Std.	bogofilter.sb	
	ham	spam
Filter	ham	6231 566
Classification	spam	0 209
Ham Misc%	0.00 (0.00-0.05)	
Spam Misc%	73.03 (69.76-76.13)	
Logit av Misc%	0.00 (0.00-nan)	
(1-ROCA)%	1.426 (0.917 - 2.211)	
Ham Misc%	Spam Misc%	
0.01	49.81 (38.14 - 61.49)	
0.02	39.23 (29.72 - 49.63)	
0.05	36.26 (27.45 - 46.09)	
0.10	30.97 (20.29 - 44.16)	
0.20	17.94 (12.93 - 24.34)	
0.50	10.19 (7.01 - 14.61)	
1.00	5.42 (3.83 - 7.62)	
2.00	2.84 (1.90 - 4.23)	
5.00	2.71 (1.81 - 4.03)	
10.00	2.71 (1.79 - 4.09)	
20.00	2.19 (1.37 - 3.51)	
50.00	1.16 (0.59 - 2.27)	
Spam Misc%	Ham Misc%	
0.01	100.00 (nan - nan)	
0.02	100.00 (nan - nan)	
0.05	100.00 (nan - nan)	
0.10	99.17 (nan - nan)	
0.20	97.64 (78.37 - 99.79)	
0.50	84.64 (53.68 - 96.32)	
1.00	55.72 (26.89 - 81.15)	
2.00	22.45 (3.94 - 67.13)	
5.00	1.12 (0.70 - 1.80)	
10.00	0.53 (0.37 - 0.76)	
20.00	0.16 (0.09 - 0.29)	
50.00	0.01 (0.00 - 0.02)	



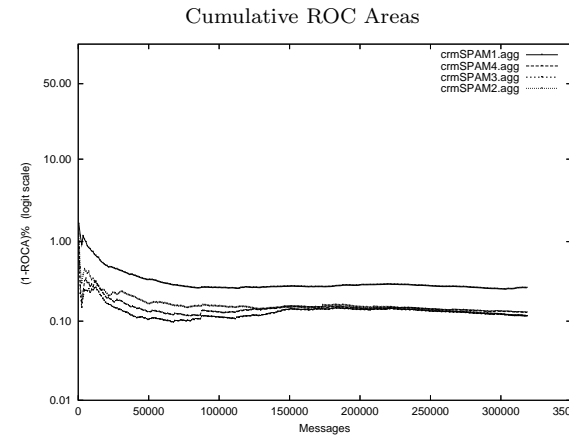
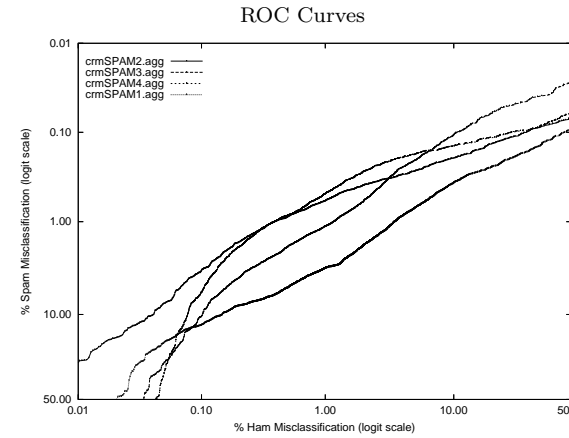
bogofilter - Bogofilter (non-participant) on Private T.M. Corpus

Gold Std.	bogofilter.tm	
Filter	ham	spam
Classification	150515	3593
	spam	170
		15923
Ham Misc%	0.11 (0.10-0.13)	
Spam Misc%	18.41 (17.87-18.96)	
Logit av Misc%	1.57 (1.46-1.70)	
(1-ROCA)%	0.792 (0.709 - 0.884)	
Ham Misc%	Spam Misc%	
0.01	82.16 (73.20 - 88.60)	
0.02	68.77 (59.53 - 76.73)	
0.05	33.67 (24.14 - 44.76)	
0.10	19.86 (18.28 - 21.53)	
0.20	13.49 (12.42 - 14.63)	
0.50	7.06 (6.45 - 7.71)	
1.00	4.78 (4.45 - 5.13)	
2.00	3.29 (3.02 - 3.58)	
5.00	1.24 (1.09 - 1.41)	
10.00	1.14 (1.00 - 1.30)	
20.00	1.00 (0.88 - 1.15)	
50.00	0.66 (0.56 - 0.76)	
Spam Misc%	Ham Misc%	
0.01	99.41 (96.99 - 99.89)	
0.02	98.41 (96.05 - 99.37)	
0.05	96.12 (93.19 - 97.82)	
0.10	91.82 (87.16 - 94.89)	
0.20	82.51 (75.62 - 87.76)	
0.50	60.55 (52.71 - 67.88)	
1.00	20.58 (12.14 - 32.71)	
2.00	3.39 (3.11 - 3.69)	
5.00	0.92 (0.81 - 1.04)	
10.00	0.34 (0.30 - 0.38)	
20.00	0.10 (0.08 - 0.11)	
50.00	0.04 (0.03 - 0.05)	

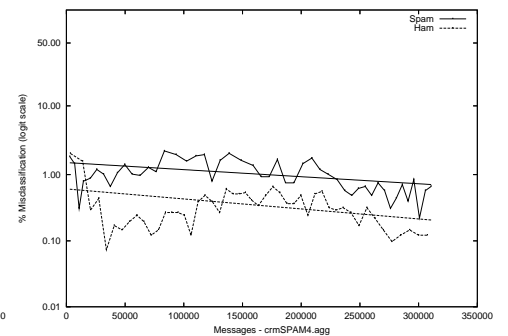
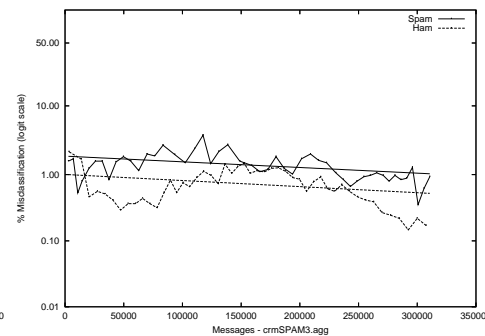
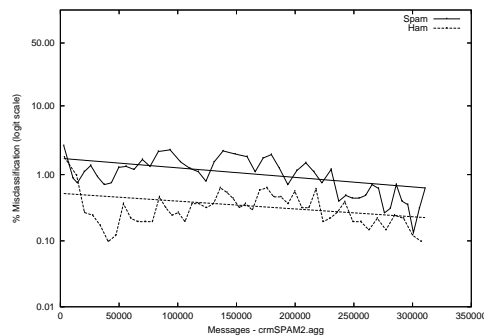
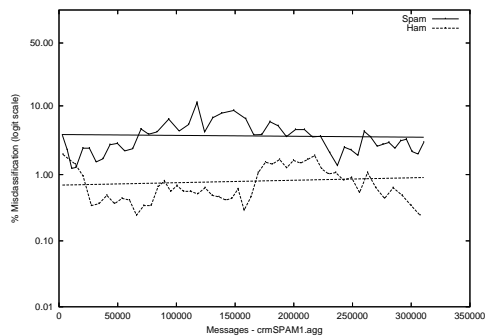


crm - Mitsubishi Electric Research Labs on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	crmSPAM1.agg		crmSPAM2.agg		crmSPAM3.agg		crmSPAM4.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.80 (0.76-0.83)		0.35 (0.33-0.38)		0.73 (0.70-0.77)		0.37 (0.34-0.40)	
Spam Misc%	3.75 (3.64-3.86)		1.08 (1.02-1.14)		1.40 (1.33-1.47)		1.05 (0.99-1.11)	
Logit av Misc%	1.74 (1.69-1.79)		0.62 (0.59-0.65)		1.01 (0.98-1.05)		0.62 (0.59-0.65)	
(1-ROCA)%	0.263 (0.246 - 0.281)		0.115 (0.103 - 0.129)		0.116 (0.106 - 0.127)		0.128 (0.114 - 0.145)	
Ham Misc%	Spam Misc%							
0.01	87.71 (73.68 - 94.79)		29.30 (20.07 - 40.62)		86.17 (82.00 - 89.49)		91.26 (87.54 - 93.94)	
0.02	50.92 (35.23 - 66.43)		16.95 (14.28 - 20.00)		76.49 (67.75 - 83.44)		81.35 (72.68 - 87.74)	
0.05	18.72 (16.25 - 21.47)		7.70 (6.41 - 9.23)		27.93 (22.01 - 34.73)		33.77 (18.20 - 53.89)	
0.10	12.79 (11.84 - 13.80)		3.46 (3.05 - 3.92)		10.50 (8.41 - 13.04)		5.90 (4.52 - 7.66)	
0.20	8.17 (7.91 - 8.43)		1.69 (1.56 - 1.83)		4.20 (3.80 - 4.65)		1.93 (1.75 - 2.13)	
0.50	5.30 (5.03 - 5.59)		0.88 (0.82 - 0.95)		1.91 (1.79 - 2.03)		0.86 (0.80 - 0.92)	
1.00	3.26 (3.13 - 3.39)		0.59 (0.54 - 0.63)		1.13 (1.05 - 1.21)		0.49 (0.44 - 0.54)	
2.00	1.85 (1.74 - 1.97)		0.40 (0.36 - 0.44)		0.61 (0.55 - 0.66)		0.28 (0.25 - 0.32)	
5.00	0.72 (0.67 - 0.78)		0.27 (0.25 - 0.30)		0.21 (0.18 - 0.24)		0.18 (0.15 - 0.21)	
10.00	0.37 (0.34 - 0.41)		0.20 (0.17 - 0.22)		0.11 (0.09 - 0.13)		0.14 (0.12 - 0.17)	
20.00	0.24 (0.21 - 0.27)		0.13 (0.11 - 0.15)		0.06 (0.04 - 0.07)		0.11 (0.10 - 0.13)	
50.00	0.09 (0.08 - 0.11)		0.07 (0.06 - 0.08)		0.03 (0.02 - 0.04)		0.06 (0.05 - 0.08)	
Spam Misc%	Ham Misc%							
0.01	99.53 (98.86 - 99.81)		97.73 (93.35 - 99.25)		92.87 (80.48 - 97.63)		97.97 (96.04 - 98.97)	
0.02	97.73 (91.53 - 99.42)		94.06 (90.95 - 96.15)		66.48 (40.04 - 85.49)		91.28 (78.59 - 96.76)	
0.05	66.26 (57.20 - 74.26)		74.10 (49.95 - 89.13)		25.18 (16.32 - 36.75)		62.28 (48.64 - 74.22)	
0.10	46.96 (39.10 - 54.97)		30.30 (23.91 - 37.56)		10.69 (8.44 - 13.46)		24.71 (15.41 - 37.15)	
0.20	23.96 (20.19 - 28.19)		9.41 (7.03 - 12.48)		5.33 (4.58 - 6.21)		3.59 (2.59 - 4.98)	
0.50	7.28 (6.76 - 7.85)		1.26 (1.11 - 1.43)		2.33 (2.17 - 2.51)		0.98 (0.89 - 1.08)	
1.00	3.57 (3.36 - 3.79)		0.39 (0.36 - 0.44)		1.17 (1.08 - 1.27)		0.40 (0.36 - 0.44)	
2.00	1.86 (1.76 - 1.97)		0.17 (0.15 - 0.19)		0.47 (0.44 - 0.50)		0.19 (0.17 - 0.21)	
5.00	0.53 (0.50 - 0.57)		0.07 (0.06 - 0.08)		0.17 (0.15 - 0.19)		0.11 (0.09 - 0.12)	
10.00	0.15 (0.13 - 0.16)		0.04 (0.03 - 0.05)		0.10 (0.09 - 0.11)		0.08 (0.07 - 0.09)	
20.00	0.05 (0.04 - 0.06)		0.02 (0.01 - 0.02)		0.06 (0.06 - 0.08)		0.06 (0.05 - 0.07)	
50.00	0.02 (0.02 - 0.03)		0.00 (0.00 - 0.01)		0.03 (0.03 - 0.04)		0.04 (0.03 - 0.05)	

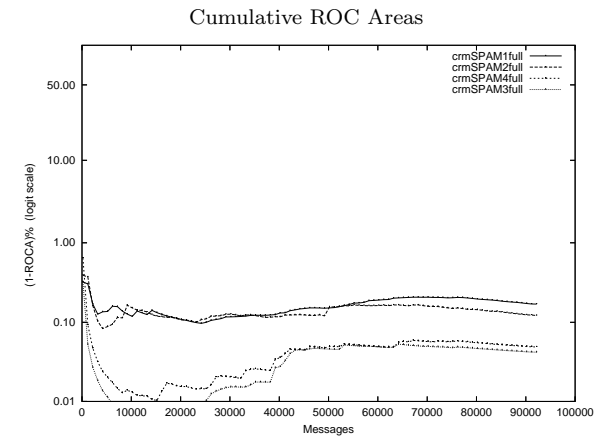
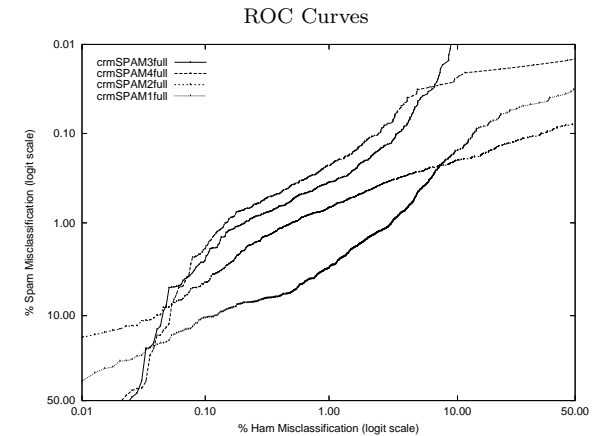


Learning Curves

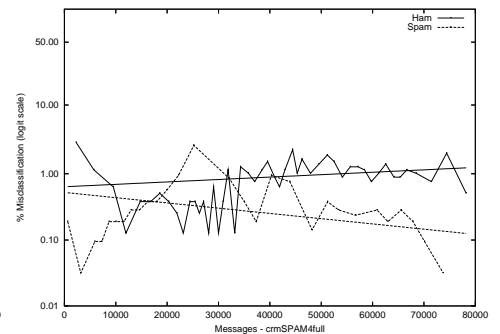
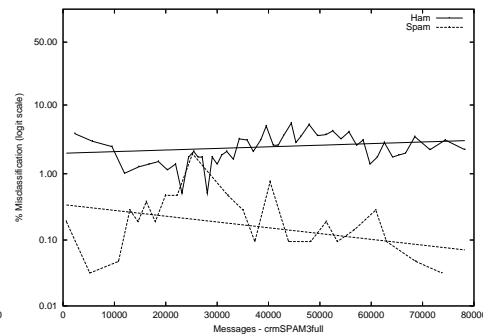
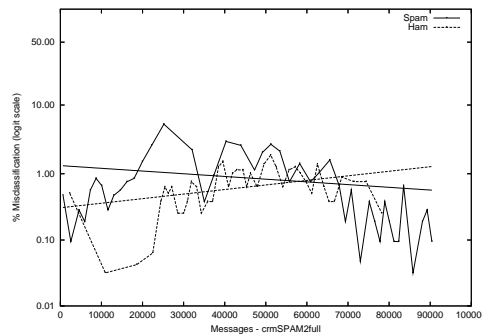
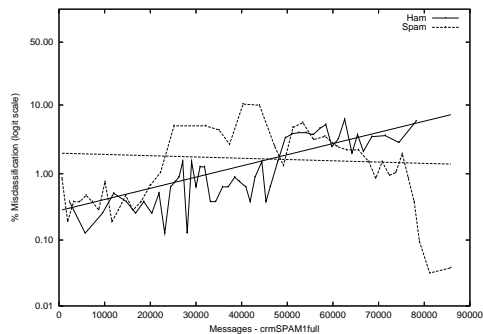


crm - Mitsubishi Electric Research Labs on Public trec05p-1/full Corpus

Gold Std. Filter Classification	crmSPAM1full		crmSPAM2full		crmSPAM3full		crmSPAM4full		
	ham	spam	ham	spam	ham	spam	ham	spam	
Filter	ham	38674	871	ham	39154	457	ham	38392	80
Classification	spam	725	51919	spam	245	52333	spam	1007	52710
Ham Misc%	1.84 (1.71-1.98)		0.62 (0.55-0.70)		2.56 (2.40-2.72)		0.91 (0.82-1.01)		
Spam Misc%	1.65 (1.54-1.76)		0.87 (0.79-0.95)		0.15 (0.12-0.19)		0.25 (0.21-0.29)		
Logit av Misc%	1.74 (1.66-1.83)		0.73 (0.68-0.79)		0.63 (0.56-0.70)		0.47 (0.43-0.52)		
(1-ROCA)%	0.169 (0.151 - 0.190)		0.122 (0.102 - 0.145)		0.042 (0.031 - 0.056)		0.049 (0.035 - 0.068)		
Ham Misc%	Spam Misc%								
0.01	68.46 (10.00 - 97.70)		17.17 (4.79 - 46.06)		93.35 (77.01 - 98.33)		94.74 (79.18 - 98.84)		
0.02	29.54 (15.29 - 49.32)		14.02 (10.66 - 18.23)		67.67 (31.83 - 90.37)		56.14 (11.11 - 92.91)		
0.05	17.98 (13.26 - 23.93)		8.25 (5.52 - 12.16)		7.88 (2.05 - 25.90)		13.06 (3.12 - 41.19)		
0.10	10.53 (8.59 - 12.84)		4.52 (3.40 - 5.98)		2.63 (1.57 - 4.39)		1.96 (1.00 - 3.80)		
0.20	7.52 (6.78 - 8.33)		2.07 (1.61 - 2.66)		1.04 (0.85 - 1.27)		0.73 (0.59 - 0.91)		
0.50	5.74 (5.03 - 6.53)		1.02 (0.90 - 1.16)		0.57 (0.48 - 0.68)		0.38 (0.32 - 0.46)		
1.00	3.14 (2.81 - 3.50)		0.68 (0.59 - 0.79)		0.35 (0.30 - 0.42)		0.23 (0.19 - 0.27)		
2.00	1.52 (1.37 - 1.70)		0.44 (0.39 - 0.51)		0.20 (0.17 - 0.24)		0.12 (0.10 - 0.15)		
5.00	0.47 (0.40 - 0.55)		0.27 (0.23 - 0.33)		0.05 (0.03 - 0.08)		0.03 (0.02 - 0.05)		
10.00	0.16 (0.13 - 0.19)		0.20 (0.16 - 0.25)		0.01 (0.00 - 0.03)		0.02 (0.01 - 0.04)		
20.00	0.06 (0.04 - 0.09)		0.14 (0.11 - 0.18)		0.00 (0.00 - 0.01)		0.02 (0.01 - 0.03)		
50.00	0.03 (0.02 - 0.05)		0.08 (0.06 - 0.11)		0.00 (0.00 - 0.00)		0.02 (0.01 - 0.03)		
Spam Misc%	Ham Misc%								
0.01	98.98 (nan - nan)		98.18 (94.99 - 99.35)		8.97 (6.68 - 11.95)		61.13 (32.20 - 83.89)		
0.02	73.14 (16.48 - 97.41)		93.52 (74.67 - 98.61)		7.47 (5.97 - 9.33)		19.86 (2.77 - 68.33)		
0.05	27.44 (13.86 - 47.05)		71.48 (61.52 - 79.71)		4.92 (4.16 - 5.82)		3.75 (2.81 - 4.98)		
0.10	13.27 (11.23 - 15.62)		36.72 (23.02 - 52.96)		3.56 (2.99 - 4.23)		2.58 (2.02 - 3.29)		
0.20	8.20 (7.09 - 9.45)		10.13 (5.90 - 16.86)		2.05 (1.67 - 2.52)		1.19 (0.92 - 1.54)		
0.50	4.90 (4.60 - 5.22)		1.69 (1.34 - 2.12)		0.57 (0.44 - 0.73)		0.36 (0.27 - 0.48)		
1.00	3.15 (2.94 - 3.38)		0.51 (0.42 - 0.62)		0.21 (0.16 - 0.27)		0.15 (0.11 - 0.20)		
2.00	1.49 (1.34 - 1.65)		0.21 (0.17 - 0.26)		0.11 (0.08 - 0.15)		0.10 (0.07 - 0.14)		
5.00	0.57 (0.49 - 0.66)		0.09 (0.06 - 0.13)		0.06 (0.04 - 0.09)		0.07 (0.04 - 0.11)		
10.00	0.12 (0.09 - 0.17)		0.04 (0.02 - 0.07)		0.05 (0.03 - 0.07)		0.05 (0.03 - 0.09)		
20.00	0.04 (0.02 - 0.07)		0.01 (0.00 - 0.03)		0.04 (0.02 - 0.06)		0.04 (0.02 - 0.07)		
50.00	0.01 (0.00 - 0.03)		0.00 (0.00 - 0.01)		0.03 (0.01 - 0.04)		0.02 (0.01 - 0.04)		

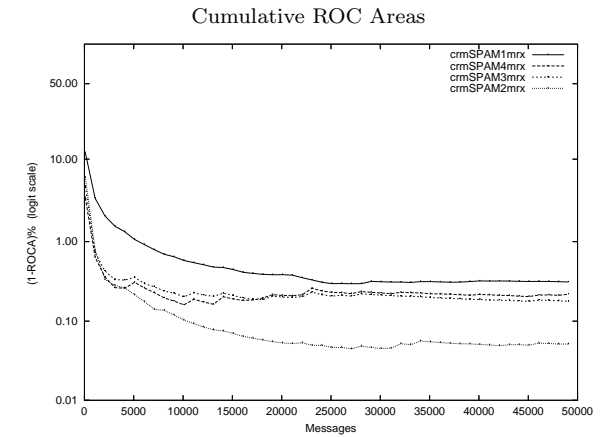
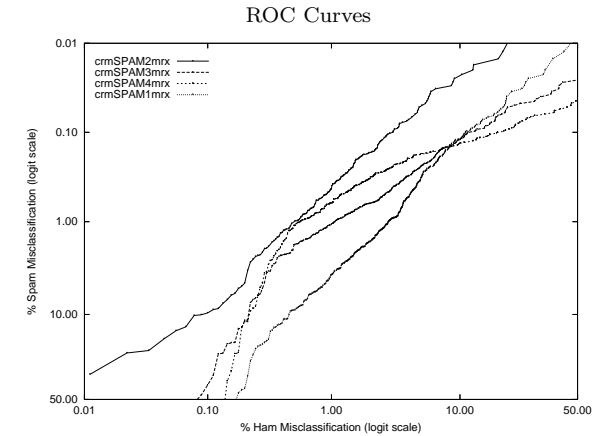


Learning Curves

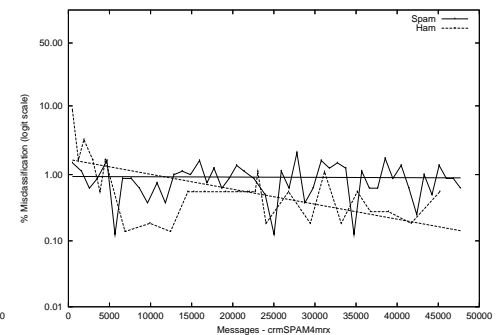
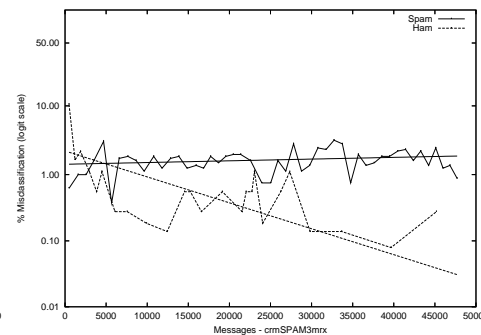
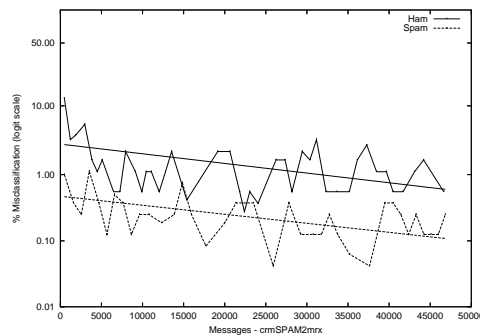
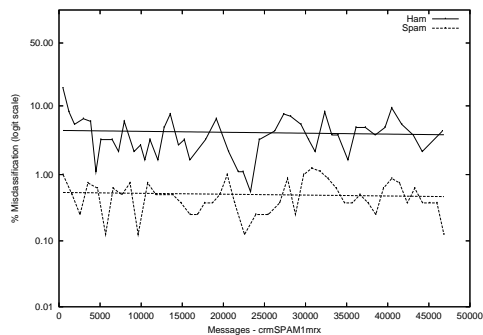


crm - Mitsubishi Electric Research Labs on Private Mr X. Corpus

Gold Std. Filter Classification	crmSPAM1mrx		crmSPAM2mrx		crmSPAM3mrx		crmSPAM4mrx	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	4.22 (3.81-4.65)		1.50 (1.26-1.78)		0.58 (0.43-0.75)		0.67 (0.52-0.87)	
Spam Misc%	0.50 (0.43-0.57)		0.24 (0.19-0.29)		1.66 (1.54-1.79)		0.91 (0.82-1.01)	
Logit av Misc%	1.46 (1.34-1.59)		0.60 (0.53-0.68)		0.98 (0.85-1.13)		0.79 (0.69-0.90)	
(1-ROCA)%	0.311 (0.244 - 0.397)		0.051 (0.035 - 0.075)		0.177 (0.128 - 0.246)		0.218 (0.157 - 0.304)	
Ham Misc%	Spam Misc%							
0.01	100.00 (nan - nan)	47.76 (30.06 - 66.04)	98.88 (93.70 - 99.81)	99.27 (97.84 - 99.75)				
0.02	99.98 (nan - nan)	34.35 (16.09 - 58.81)	94.51 (73.36 - 99.08)	98.77 (95.08 - 99.70)				
0.05	97.39 (46.76 - 99.94)	17.24 (8.51 - 31.81)	85.08 (38.99 - 98.07)	95.55 (86.41 - 98.64)				
0.10	81.61 (53.16 - 94.55)	9.65 (5.45 - 16.53)	40.82 (11.62 - 78.35)	82.36 (30.33 - 98.04)				
0.20	43.00 (17.58 - 72.73)	4.71 (1.90 - 11.20)	11.51 (3.69 - 30.65)	11.80 (1.63 - 51.88)				
0.50	9.07 (6.61 - 12.33)	1.05 (0.75 - 1.47)	1.81 (1.29 - 2.53)	1.11 (0.70 - 1.75)				
1.00	3.85 (2.82 - 5.22)	0.43 (0.31 - 0.62)	1.07 (0.86 - 1.35)	0.63 (0.50 - 0.79)				
2.00	1.68 (1.38 - 2.04)	0.17 (0.13 - 0.22)	0.63 (0.55 - 0.73)	0.33 (0.28 - 0.41)				
5.00	0.36 (0.28 - 0.46)	0.05 (0.03 - 0.08)	0.25 (0.20 - 0.31)	0.18 (0.14 - 0.23)				
10.00	0.12 (0.09 - 0.16)	0.02 (0.01 - 0.05)	0.12 (0.09 - 0.17)	0.13 (0.10 - 0.18)				
20.00	0.05 (0.03 - 0.08)	0.01 (0.00 - 0.04)	0.06 (0.04 - 0.09)	0.10 (0.07 - 0.13)				
50.00	0.01 (0.00 - 0.03)	0.01 (0.00 - 0.03)	0.03 (0.02 - 0.05)	0.04 (0.03 - 0.07)				
Spam Misc%	Ham Misc%							
0.01	46.91 (31.50 - 62.93)	21.22 (3.91 - 64.08)	86.72 (71.73 - 94.39)	86.87 (82.23 - 90.44)				
0.02	35.27 (21.71 - 51.72)	12.39 (4.55 - 29.57)	76.64 (38.75 - 94.45)	81.72 (63.88 - 91.87)				
0.05	20.34 (15.43 - 26.32)	5.17 (3.82 - 6.96)	23.48 (13.05 - 38.54)	46.89 (28.73 - 65.91)				
0.10	10.96 (8.27 - 14.40)	3.09 (2.31 - 4.11)	11.46 (7.73 - 16.67)	18.31 (11.29 - 28.30)				
0.20	6.92 (5.96 - 8.02)	1.63 (1.26 - 2.11)	6.36 (5.15 - 7.84)	4.03 (2.79 - 5.78)				
0.50	4.22 (3.75 - 4.73)	0.91 (0.70 - 1.17)	2.67 (2.22 - 3.20)	1.19 (0.96 - 1.49)				
1.00	2.85 (2.49 - 3.27)	0.51 (0.37 - 0.69)	1.07 (0.83 - 1.39)	0.56 (0.42 - 0.75)				
2.00	1.70 (1.40 - 2.07)	0.29 (0.19 - 0.43)	0.50 (0.37 - 0.68)	0.39 (0.26 - 0.57)				
5.00	0.87 (0.68 - 1.13)	0.20 (0.12 - 0.33)	0.29 (0.19 - 0.44)	0.28 (0.18 - 0.42)				
10.00	0.46 (0.34 - 0.64)	0.10 (0.04 - 0.23)	0.22 (0.13 - 0.36)	0.22 (0.14 - 0.35)				
20.00	0.27 (0.18 - 0.39)	0.04 (0.01 - 0.15)	0.14 (0.08 - 0.25)	0.19 (0.11 - 0.31)				
50.00	0.18 (0.10 - 0.30)	0.01 (0.01 - 0.01)	0.09 (0.04 - 0.19)	0.14 (0.09 - 0.24)				

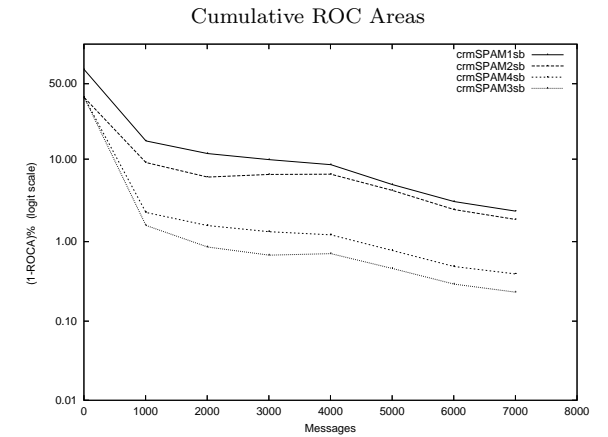
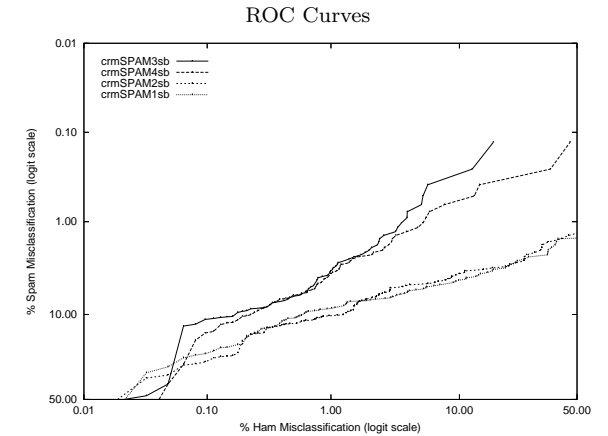


Learning Curves

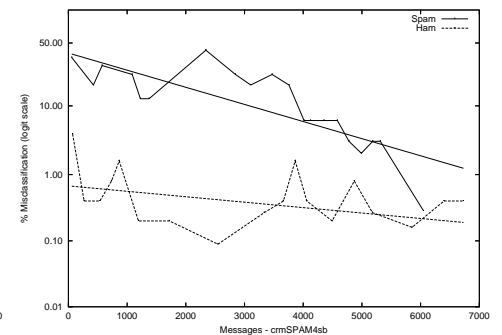
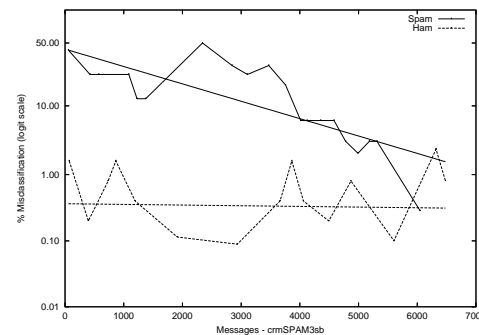
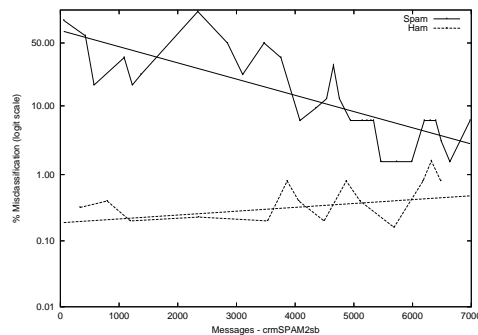
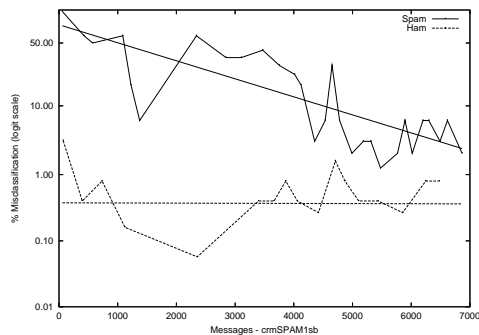


crm - Mitsubishi Electric Research Labs on Private S.B. Corpus

Gold Std. Filter Classification	crmSPAM1sb			crmSPAM2sb			crmSPAM3sb			crmSPAM4sb		
	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	spam ham	
Ham Misc%	0.37 (0.23-0.55)			0.30 (0.18-0.48)			0.34 (0.21-0.51)			0.39 (0.25-0.57)		
Spam Misc%	13.42 (11.10-16.02)			13.55 (11.22-16.16)			7.61 (5.85-9.71)			6.97 (5.28-8.99)		
Logit av Misc%	2.34 (1.87-2.93)			2.14 (1.68-2.73)			1.64 (1.28-2.10)			1.67 (1.32-2.12)		
(1-ROCA)%	2.393 (1.689 - 3.382)			1.888 (1.315 - 2.704)			0.231 (0.142 - 0.377)			0.393 (0.203 - 0.759)		
Ham Misc%										Spam Misc%		
0.01	91.48 (57.06 - 98.86)			74.58 (47.94 - 90.34)			92.77 (48.17 - 99.44)			93.94 (68.00 - 99.12)		
0.02	62.32 (12.76 - 94.92)			54.06 (21.06 - 83.85)			51.61 (7.75 - 93.13)			84.65 (27.17 - 98.79)		
0.05	30.06 (15.02 - 51.11)			34.71 (17.14 - 57.73)			40.13 (8.57 - 82.74)			40.13 (5.56 - 88.42)		
0.10	23.48 (17.49 - 30.77)			27.48 (18.04 - 39.49)			11.23 (4.38 - 25.87)			15.23 (4.26 - 42.06)		
0.20	19.35 (13.52 - 26.93)			18.06 (10.85 - 28.54)			9.29 (7.04 - 12.16)			10.45 (7.03 - 15.26)		
0.50	10.97 (8.54 - 13.98)			12.39 (10.13 - 15.07)			6.71 (4.79 - 9.33)			6.32 (4.60 - 8.63)		
1.00	8.77 (7.05 - 10.88)			10.32 (8.32 - 12.74)			3.48 (2.26 - 5.32)			3.74 (2.32 - 5.97)		
2.00	7.23 (5.54 - 9.38)			6.97 (5.40 - 8.95)			2.19 (1.33 - 3.59)			2.45 (1.39 - 4.30)		
5.00	5.55 (4.17 - 7.36)			4.90 (3.53 - 6.76)			0.77 (0.27 - 2.21)			1.16 (0.54 - 2.47)		
10.00	4.39 (3.20 - 5.99)			3.74 (2.62 - 5.33)			0.39 (0.11 - 1.40)			0.65 (0.22 - 1.91)		
20.00	3.35 (2.29 - 4.89)			3.23 (2.10 - 4.92)			0.13 (0.04 - 0.44)			0.39 (0.11 - 1.40)		
50.00	1.55 (0.91 - 2.61)			1.42 (0.75 - 2.67)			0.06 (0.06 - 0.07)			0.13 (0.04 - 0.42)		
Spam Misc%										Ham Misc%		
0.01	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.10	99.95 (nan - nan)			85.33 (80.36 - 89.21)			36.49 (13.89 - 67.19)			63.99 (33.61 - 86.18)		
0.20	99.76 (98.29 - 99.97)			81.09 (73.57 - 86.86)			17.32 (3.73 - 53.10)			46.77 (11.75 - 85.28)		
0.50	99.60 (97.94 - 99.92)			76.46 (63.02 - 86.09)			5.79 (1.84 - 16.81)			13.93 (2.25 - 53.28)		
1.00	97.11 (57.03 - 99.88)			62.35 (37.84 - 81.84)			4.01 (2.40 - 6.63)			5.62 (1.99 - 14.86)		
2.00	38.24 (10.39 - 76.79)			34.09 (17.76 - 55.32)			2.10 (1.21 - 3.62)			2.66 (1.22 - 5.73)		
5.00	7.16 (2.68 - 17.73)			4.20 (1.59 - 10.63)			0.77 (0.49 - 1.22)			0.77 (0.47 - 1.26)		
10.00	0.61 (0.32 - 1.16)			1.32 (0.66 - 2.60)			0.18 (0.08 - 0.40)			0.24 (0.13 - 0.46)		
20.00	0.18 (0.08 - 0.39)			0.19 (0.10 - 0.38)			0.06 (0.02 - 0.20)			0.08 (0.03 - 0.25)		
50.00	0.03 (0.01 - 0.10)			0.03 (0.01 - 0.13)			0.03 (0.01 - 0.13)			0.05 (0.01 - 0.19)		

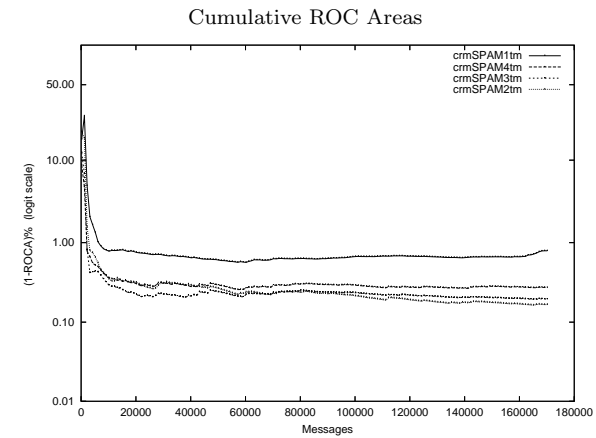
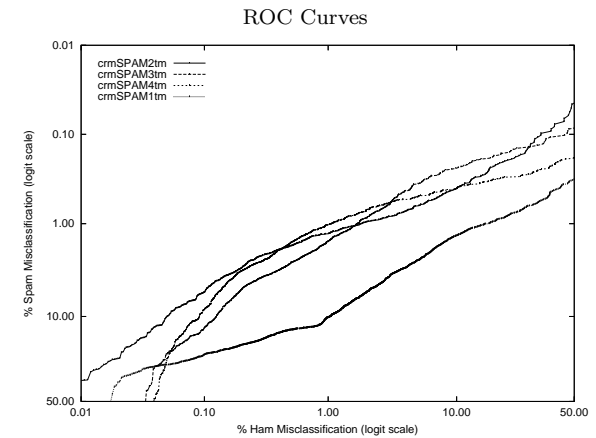


Learning Curves

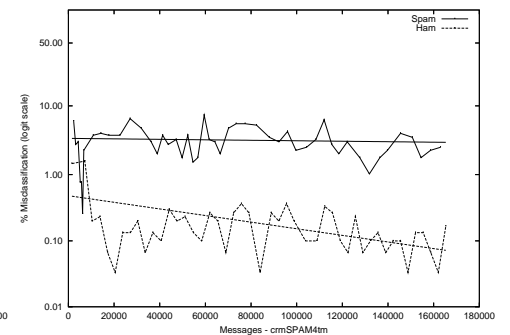
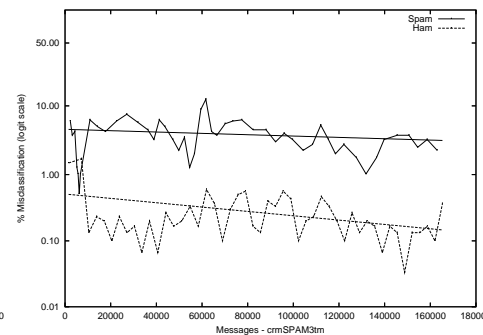
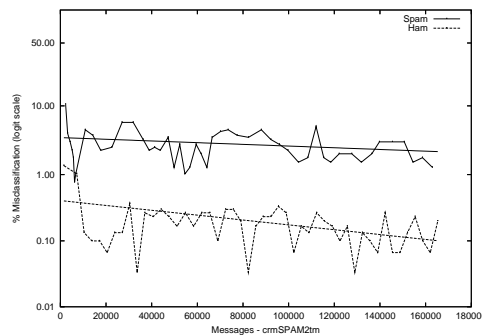
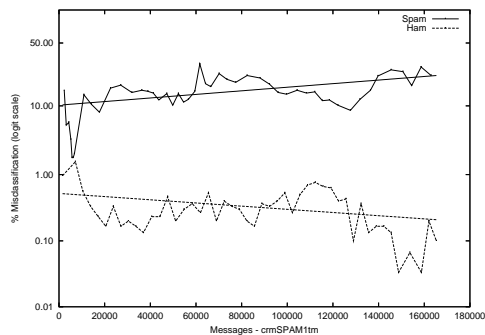


crm - Mitsubishi Electric Research Labs on Private T.M. Corpus

Gold Std. Filter Classification	crmSPAM1tm		crmSPAM2tm		crmSPAM3tm		crmSPAM4tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.33 (0.31-0.36)		0.21 (0.19-0.24)		0.28 (0.26-0.31)		0.21 (0.18-0.23)	
Spam Misc%	15.72 (15.21-16.24)		2.91 (2.67-3.15)		3.99 (3.72-4.28)		3.26 (3.01-3.52)	
Logit av Misc%	2.44 (2.33-2.56)		0.79 (0.74-0.85)		1.08 (1.01-1.14)		0.83 (0.77-0.89)	
(1-ROCA)%	0.790 (0.720 - 0.868)		0.166 (0.138 - 0.201)		0.195 (0.160 - 0.238)		0.272 (0.225 - 0.329)	
Ham Misc%	Spam Misc%							
0.01	65.45 (40.59 - 84.01)	36.77 (25.23 - 50.05)	87.15 (82.63 - 90.63)	86.51 (80.70 - 90.77)	65.45 (40.59 - 84.01)	36.77 (25.23 - 50.05)	87.15 (82.63 - 90.63)	86.51 (80.70 - 90.77)
0.02	37.96 (25.99 - 51.61)	24.85 (18.91 - 31.93)	77.90 (63.06 - 87.92)	77.58 (71.27 - 82.83)	37.96 (25.99 - 51.61)	24.85 (18.91 - 31.93)	77.90 (63.06 - 87.92)	77.58 (71.27 - 82.83)
0.05	28.03 (26.82 - 29.28)	10.03 (7.60 - 13.13)	22.81 (16.55 - 30.58)	25.04 (12.49 - 43.87)	28.03 (26.82 - 29.28)	10.03 (7.60 - 13.13)	22.81 (16.55 - 30.58)	25.04 (12.49 - 43.87)
0.10	23.12 (21.61 - 24.70)	5.64 (4.71 - 6.74)	13.06 (10.64 - 15.94)	8.59 (6.71 - 10.95)	23.12 (21.61 - 24.70)	5.64 (4.71 - 6.74)	13.06 (10.64 - 15.94)	8.59 (6.71 - 10.95)
0.20	19.01 (18.04 - 20.02)	3.09 (2.71 - 3.52)	5.30 (4.45 - 6.30)	3.34 (2.94 - 3.78)	19.01 (18.04 - 20.02)	3.09 (2.71 - 3.52)	5.30 (4.45 - 6.30)	3.34 (2.94 - 3.78)
0.50	13.62 (13.05 - 14.21)	1.77 (1.58 - 1.98)	2.74 (2.43 - 3.08)	1.64 (1.46 - 1.84)	13.62 (13.05 - 14.21)	1.77 (1.58 - 1.98)	2.74 (2.43 - 3.08)	1.64 (1.46 - 1.84)
1.00	10.28 (9.71 - 10.89)	1.30 (1.15 - 1.46)	1.58 (1.40 - 1.78)	1.04 (0.92 - 1.16)	10.28 (9.71 - 10.89)	1.30 (1.15 - 1.46)	1.58 (1.40 - 1.78)	1.04 (0.92 - 1.16)
2.00	5.73 (5.33 - 6.16)	0.96 (0.85 - 1.09)	0.87 (0.75 - 1.01)	0.73 (0.62 - 0.86)	5.73 (5.33 - 6.16)	0.96 (0.85 - 1.09)	0.87 (0.75 - 1.01)	0.73 (0.62 - 0.86)
5.00	2.60 (2.38 - 2.84)	0.64 (0.53 - 0.76)	0.36 (0.29 - 0.46)	0.51 (0.42 - 0.62)	2.60 (2.38 - 2.84)	0.64 (0.53 - 0.76)	0.36 (0.29 - 0.46)	0.51 (0.42 - 0.62)
10.00	1.36 (1.21 - 1.53)	0.41 (0.32 - 0.52)	0.24 (0.18 - 0.32)	0.41 (0.33 - 0.51)	1.36 (1.21 - 1.53)	0.41 (0.32 - 0.52)	0.24 (0.18 - 0.32)	0.41 (0.33 - 0.51)
20.00	0.89 (0.77 - 1.02)	0.21 (0.15 - 0.30)	0.17 (0.12 - 0.25)	0.32 (0.25 - 0.40)	0.89 (0.77 - 1.02)	0.21 (0.15 - 0.30)	0.17 (0.12 - 0.25)	0.32 (0.25 - 0.40)
50.00	0.32 (0.25 - 0.42)	0.05 (0.02 - 0.10)	0.09 (0.05 - 0.14)	0.18 (0.13 - 0.25)	0.32 (0.25 - 0.42)	0.05 (0.02 - 0.10)	0.09 (0.05 - 0.14)	0.18 (0.13 - 0.25)
Spam Misc%	Ham Misc%							
0.01	99.93 (99.68 - 99.98)	75.17 (31.01 - 95.33)	99.92 (92.88 - 100.00)	99.97 (99.57 - 100.00)	99.93 (99.68 - 99.98)	75.17 (31.01 - 95.33)	99.92 (92.88 - 100.00)	99.97 (99.57 - 100.00)
0.02	99.59 (97.50 - 99.93)	60.54 (38.30 - 79.13)	95.11 (30.52 - 99.88)	99.54 (94.10 - 99.97)	99.59 (97.50 - 99.93)	60.54 (38.30 - 79.13)	95.11 (30.52 - 99.88)	99.54 (94.10 - 99.97)
0.05	97.96 (87.08 - 99.71)	48.91 (40.60 - 57.29)	81.53 (54.81 - 94.14)	93.61 (68.56 - 98.99)	97.96 (87.08 - 99.71)	48.91 (40.60 - 57.29)	81.53 (54.81 - 94.14)	93.61 (68.56 - 98.99)
0.10	76.49 (46.49 - 92.42)	34.94 (25.92 - 45.19)	46.74 (25.66 - 69.05)	77.19 (60.83 - 88.06)	76.49 (46.49 - 92.42)	34.94 (25.92 - 45.19)	46.74 (25.66 - 69.05)	77.19 (60.83 - 88.06)
0.20	64.11 (58.06 - 69.74)	21.45 (13.91 - 31.57)	12.97 (6.73 - 23.54)	43.50 (28.81 - 59.42)	64.11 (58.06 - 69.74)	21.45 (13.91 - 31.57)	12.97 (6.73 - 23.54)	43.50 (28.81 - 59.42)
0.50	38.13 (33.24 - 43.28)	7.48 (5.49 - 10.11)	3.55 (2.90 - 4.35)	5.20 (2.75 - 9.63)	38.13 (33.24 - 43.28)	7.48 (5.49 - 10.11)	3.55 (2.90 - 4.35)	5.20 (2.75 - 9.63)
1.00	16.93 (13.00 - 21.74)	1.86 (1.34 - 2.57)	1.71 (1.49 - 1.97)	1.07 (0.86 - 1.34)	16.93 (13.00 - 21.74)	1.86 (1.34 - 2.57)	1.71 (1.49 - 1.97)	1.07 (0.86 - 1.34)
2.00	6.57 (6.00 - 7.19)	0.38 (0.30 - 0.48)	0.78 (0.68 - 0.89)	0.39 (0.35 - 0.44)	6.57 (6.00 - 7.19)	0.38 (0.30 - 0.48)	0.78 (0.68 - 0.89)	0.39 (0.35 - 0.44)
5.00	2.35 (2.13 - 2.58)	0.11 (0.09 - 0.13)	0.21 (0.18 - 0.24)	0.14 (0.12 - 0.16)	2.35 (2.13 - 2.58)	0.11 (0.09 - 0.13)	0.21 (0.18 - 0.24)	0.14 (0.12 - 0.16)
10.00	1.05 (0.97 - 1.13)	0.05 (0.04 - 0.06)	0.13 (0.11 - 0.15)	0.09 (0.08 - 0.11)	1.05 (0.97 - 1.13)	0.05 (0.04 - 0.06)	0.13 (0.11 - 0.15)	0.09 (0.08 - 0.11)
20.00	0.17 (0.15 - 0.21)	0.02 (0.02 - 0.03)	0.06 (0.05 - 0.08)	0.06 (0.05 - 0.07)	0.17 (0.15 - 0.21)	0.02 (0.02 - 0.03)	0.06 (0.05 - 0.08)	0.06 (0.05 - 0.07)
50.00	0.02 (0.01 - 0.03)	0.01 (0.00 - 0.01)	0.03 (0.03 - 0.05)	0.04 (0.03 - 0.05)	0.02 (0.01 - 0.03)	0.01 (0.00 - 0.01)	0.03 (0.03 - 0.05)	0.04 (0.03 - 0.05)



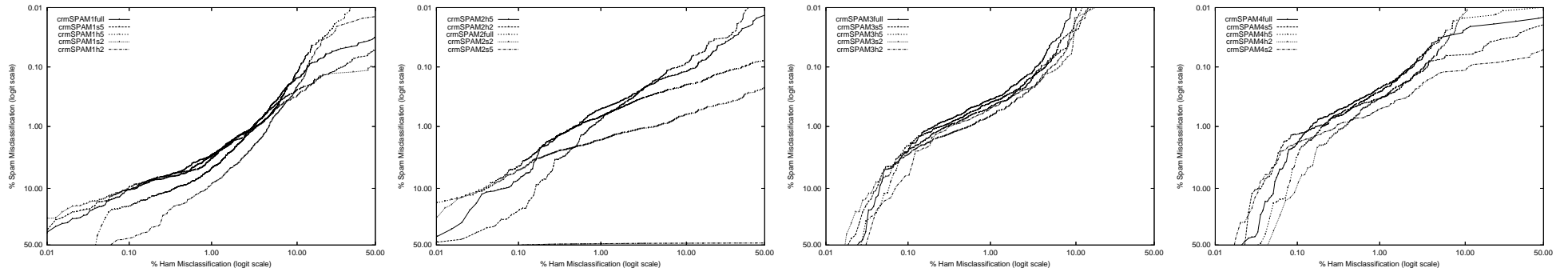
Learning Curves



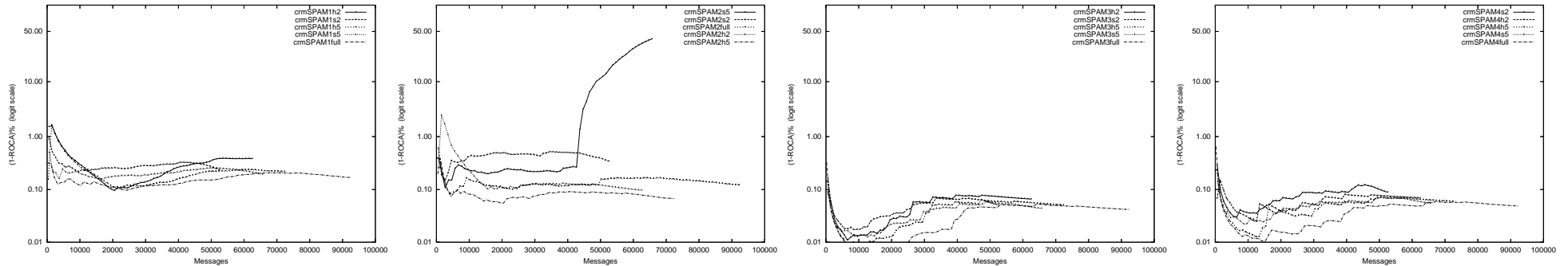
crm - Mitsubishi Electric Research Labs - Public subset Corpora

	crmSPAM1full	crmSPAM1h2	crmSPAM1h5	crmSPAM1s2	crmSPAM1s5
Ham Misc%	1.84 (1.71-1.98)	5.98 (5.52-6.47)	3.47 (3.21-3.73)	0.22 (0.18-0.28)	0.68 (0.60-0.76)
Spam Misc%	1.65 (1.54-1.76)	0.59 (0.53-0.66)	1.00 (0.92-1.09)	6.76 (6.34-7.20)	3.79 (3.56-4.02)
Logit av Misc% (1-ROCA)%	1.74 (1.66-1.83)	1.91 (1.79-2.05)	1.87 (1.77-1.98)	1.26 (1.13-1.40)	1.61 (1.51-1.72)
	0.169 (0.151 - 0.190)	0.383 (0.328 - 0.447)	0.219 (0.193 - 0.248)	0.236 (0.188 - 0.296)	0.194 (0.165 - 0.227)
	crmSPAM2full	crmSPAM2h2	crmSPAM2h5	crmSPAM2s2	crmSPAM2s5
Ham Misc%	0.62 (0.55-0.70)	2.11 (1.84-2.42)	0.97 (0.84-1.12)	0.28 (0.23-0.34)	0.27 (0.22-0.32)
Spam Misc%	0.87 (0.79-0.95)	0.38 (0.32-0.43)	0.53 (0.47-0.59)	2.67 (2.40-2.96)	49.18 (48.58-49.79)
Logit av Misc% (1-ROCA)%	0.73 (0.68-0.79)	0.89 (0.81-0.98)	0.71 (0.65-0.78)	0.87 (0.78-0.97)	4.84 (4.41-5.30)
	0.122 (0.102 - 0.145)	0.097 (0.074 - 0.127)	0.067 (0.054 - 0.083)	0.343 (0.270 - 0.435)	41.915 (41.362 - 42.470)
	crmSPAM3full	crmSPAM3h2	crmSPAM3h5	crmSPAM3s2	crmSPAM3s5
Ham Misc%	2.56 (2.40-2.72)	4.12 (3.74-4.54)	3.17 (2.92-3.42)	2.41 (2.26-2.57)	2.48 (2.33-2.64)
Spam Misc%	0.15 (0.12-0.19)	0.16 (0.13-0.20)	0.15 (0.12-0.19)	0.33 (0.24-0.44)	0.23 (0.18-0.30)
Logit av Misc% (1-ROCA)%	0.63 (0.56-0.70)	0.82 (0.73-0.92)	0.70 (0.63-0.79)	0.89 (0.77-1.04)	0.76 (0.67-0.87)
	0.042 (0.031 - 0.056)	0.066 (0.041 - 0.105)	0.051 (0.032 - 0.080)	0.051 (0.039 - 0.065)	0.044 (0.033 - 0.058)
	crmSPAM4full	crmSPAM4h2	crmSPAM4h5	crmSPAM4s2	crmSPAM4s5
Ham Misc%	0.91 (0.82-1.01)	3.56 (3.20-3.95)	1.96 (1.77-2.16)	0.61 (0.54-0.69)	0.73 (0.65-0.82)
Spam Misc%	0.25 (0.21-0.29)	0.09 (0.07-0.12)	0.13 (0.10-0.17)	0.72 (0.58-0.88)	0.40 (0.33-0.48)
Logit av Misc% (1-ROCA)%	0.47 (0.43-0.52)	0.57 (0.49-0.66)	0.51 (0.45-0.58)	0.66 (0.59-0.75)	0.54 (0.48-0.60)
	0.049 (0.035 - 0.068)	0.069 (0.041 - 0.117)	0.059 (0.038 - 0.093)	0.089 (0.061 - 0.129)	0.055 (0.039 - 0.077)

ROC Curves

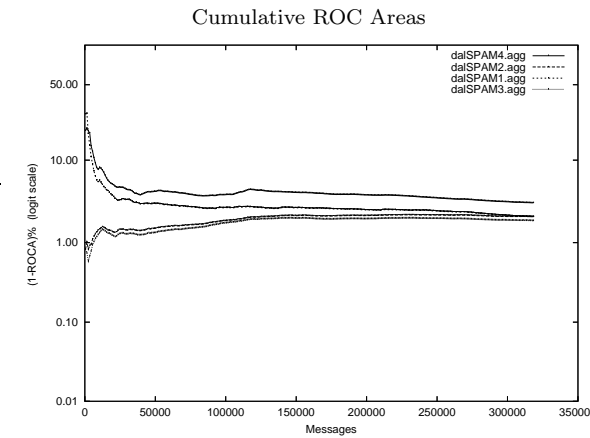
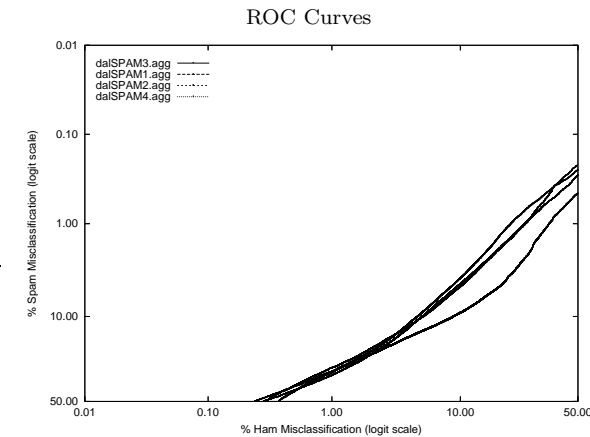


Cumulative ROC Area

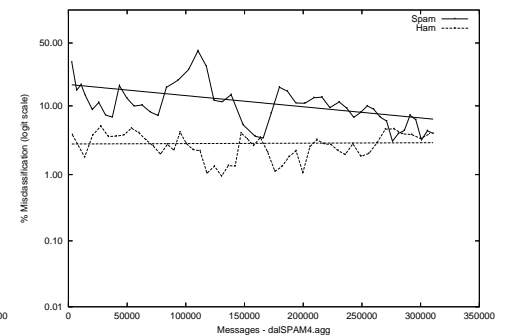
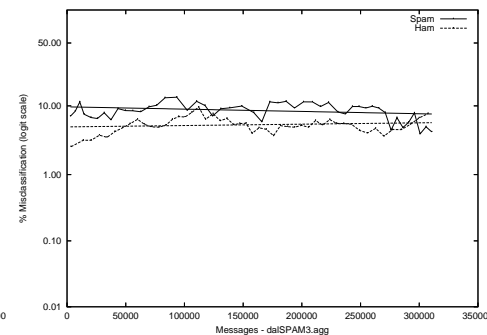
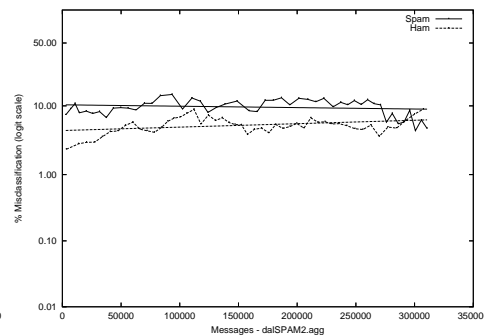
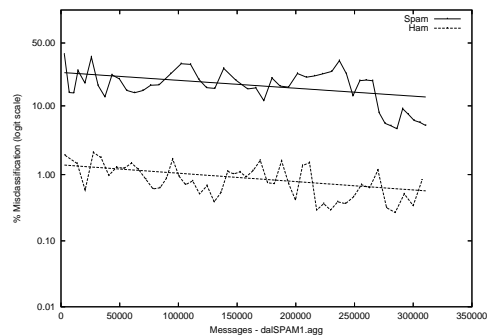


dal - Dalhousie U. on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	dalSPAM1.agg			dalSPAM2.agg			dalSPAM3.agg			dalSPAM4.agg		
	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	
Ham Misc%	0.92 (0.88-0.96)			5.40 (5.30-5.50)			5.44 (5.35-5.54)			2.92 (2.85-2.99)		
Spam Misc%	18.93 (18.70-19.16)			9.64 (9.47-9.82)			8.65 (8.48-8.81)			11.66 (11.47-11.85)		
Logit av Misc%	4.44 (4.34-4.54)			7.24 (7.15-7.33)			6.87 (6.79-6.97)			5.93 (5.84-6.01)		
(1-ROCA)%	2.097 (2.053 - 2.141)			2.100 (2.060 - 2.141)			1.873 (1.834 - 1.913)			3.115 (3.062 - 3.170)		
Ham Misc%										Spam Misc%		
0.01	99.64 (99.59 - 99.69)			78.85 (75.82 - 81.58)			79.93 (75.80 - 83.50)			90.33 (87.59 - 92.52)		
0.02	99.57 (99.51 - 99.62)			74.19 (71.97 - 76.29)			72.40 (69.77 - 74.89)			86.54 (84.74 - 88.16)		
0.05	99.42 (99.34 - 99.49)			66.35 (64.95 - 67.73)			65.71 (64.47 - 66.92)			80.52 (78.26 - 82.60)		
0.10	99.15 (98.52 - 99.52)			60.64 (59.13 - 62.13)			59.50 (57.78 - 61.18)			79.14 (78.90 - 79.37)		
0.20	65.19 (59.67 - 70.33)			54.23 (53.08 - 55.37)			51.98 (51.06 - 52.91)			73.02 (68.72 - 76.93)		
0.50	39.95 (38.96 - 40.94)			43.04 (42.26 - 43.83)			40.46 (39.55 - 41.38)			43.81 (42.55 - 45.07)		
1.00	29.66 (29.07 - 30.24)			33.92 (33.26 - 34.60)			31.74 (31.09 - 32.39)			32.34 (31.69 - 32.99)		
2.00	20.72 (20.29 - 21.17)			23.57 (23.10 - 24.05)			22.03 (21.54 - 22.53)			23.26 (22.91 - 23.63)		
5.00	10.56 (10.34 - 10.79)			10.44 (10.17 - 10.72)			9.55 (9.28 - 9.83)			14.33 (14.08 - 14.58)		
10.00	4.93 (4.79 - 5.07)			4.62 (4.46 - 4.78)			3.99 (3.87 - 4.12)			9.25 (9.07 - 9.44)		
20.00	1.74 (1.65 - 1.83)			1.69 (1.63 - 1.76)			1.22 (1.16 - 1.28)			4.49 (4.35 - 4.64)		
50.00	0.22 (0.20 - 0.25)			0.29 (0.26 - 0.32)			0.25 (0.22 - 0.29)			0.46 (0.42 - 0.51)		
Spam Misc%										Ham Misc%		
0.01	100.00 (nan - nan)			85.38 (81.29 - 88.70)			87.63 (83.81 - 90.65)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			81.54 (78.37 - 84.34)			84.54 (80.63 - 87.78)			100.00 (nan - nan)		
0.05	98.46 (nan - nan)			72.96 (70.23 - 75.52)			74.33 (71.36 - 77.09)			98.78 (nan - nan)		
0.10	70.46 (63.72 - 76.40)			65.46 (63.08 - 67.75)			65.31 (62.89 - 67.66)			83.60 (77.33 - 88.39)		
0.20	53.32 (49.48 - 57.12)			56.06 (54.10 - 58.00)			54.77 (52.76 - 56.77)			68.89 (65.56 - 72.03)		
0.50	36.42 (35.32 - 37.54)			39.73 (38.06 - 41.43)			33.87 (32.18 - 35.60)			48.28 (46.67 - 49.89)		
1.00	27.47 (26.53 - 28.44)			27.17 (26.49 - 27.86)			22.29 (21.64 - 22.96)			37.01 (36.25 - 37.78)		
2.00	18.35 (17.75 - 18.96)			18.11 (17.70 - 18.53)			15.50 (15.09 - 15.91)			29.26 (28.78 - 29.75)		
5.00	9.91 (9.70 - 10.12)			9.35 (9.13 - 9.58)			8.47 (8.26 - 8.67)			18.54 (18.06 - 19.03)		
10.00	5.34 (5.21 - 5.47)			5.22 (5.08 - 5.36)			4.79 (4.67 - 4.91)			8.97 (8.74 - 9.20)		
20.00	2.12 (2.05 - 2.18)			2.52 (2.44 - 2.60)			2.31 (2.23 - 2.39)			2.71 (2.62 - 2.80)		
50.00	0.30 (0.27 - 0.33)			0.28 (0.26 - 0.31)			0.24 (0.22 - 0.26)			0.37 (0.35 - 0.40)		

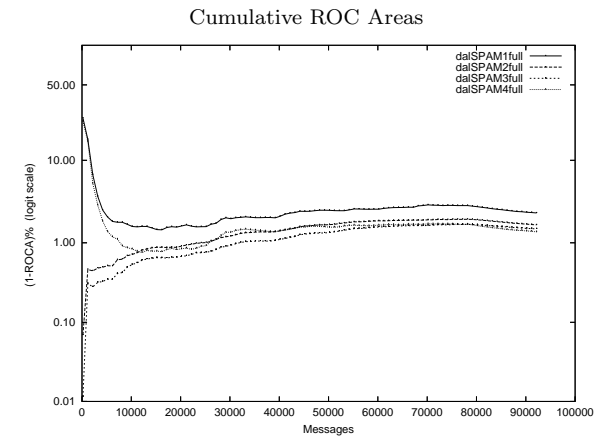
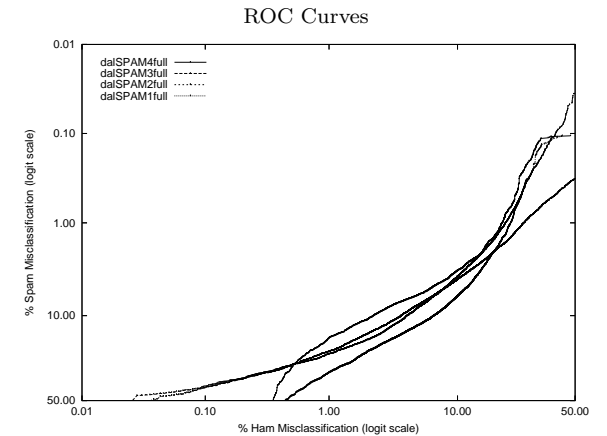


Learning Curves

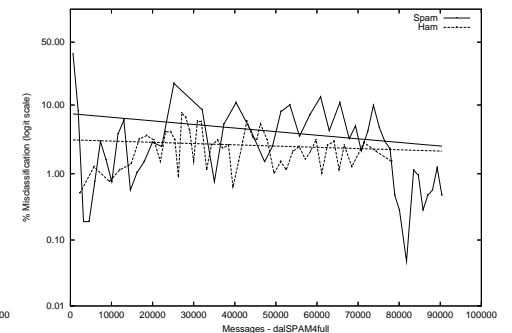
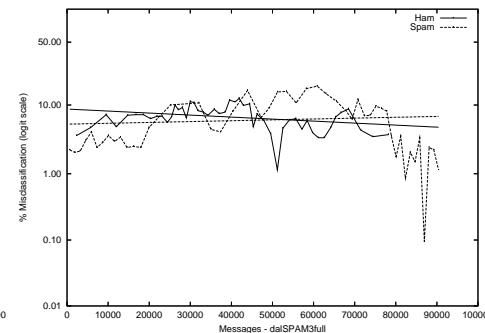
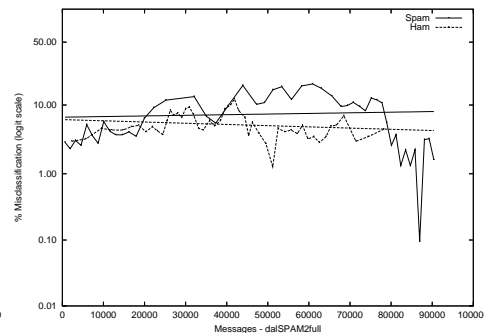
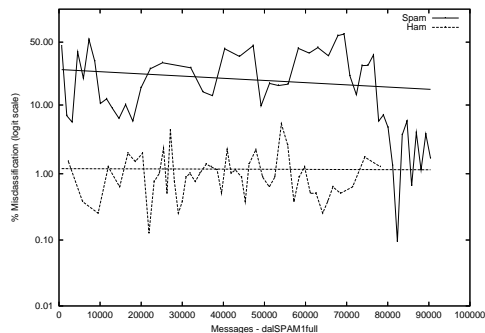


dal - Dalhousie U. on Public trec05p-1/full Corpus

Gold Std. Filter Classification	dalSPAM1full			dalSPAM2full			dalSPAM3full			dalSPAM4full		
	ham spam	ham 38937 462	spam 11125 41665	ham spam	ham 37297 2102	spam 3968 48822	ham spam	ham 36719 2680	spam 3291 49499	ham spam	ham 38338 1061	spam 2376 50414
Ham Misc%	1.17 (1.07-1.28)			5.34 (5.12-5.56)			6.80 (6.56-7.06)			2.69 (2.54-2.86)		
Spam Misc%	21.07 (20.73-21.42)			7.52 (7.29-7.74)			6.23 (6.03-6.44)			4.50 (4.33-4.68)		
Logit av Misc%	5.33 (5.10-5.57)			6.34 (6.18-6.50)			6.51 (6.35-6.67)			3.49 (3.36-3.61)		
(1-ROCA)%	2.348 (2.255 - 2.444)			1.674 (1.609 - 1.742)			1.491 (1.441 - 1.544)			1.370 (1.306 - 1.436)		
Ham Misc%	Spam Misc%											
0.01	99.97 (99.93 - 99.99)			61.64 (51.41 - 70.94)			59.34 (50.69 - 67.44)			76.80 (75.19 - 78.34)		
0.02	99.93 (99.87 - 99.97)			55.70 (49.32 - 61.90)			53.86 (47.88 - 59.73)			76.77 (76.37 - 77.17)		
0.05	99.87 (99.77 - 99.92)			47.22 (43.79 - 50.68)			45.10 (43.31 - 46.91)			76.71 (76.30 - 77.10)		
0.10	99.75 (99.65 - 99.83)			41.92 (39.17 - 44.72)			41.00 (38.51 - 43.53)			76.58 (76.18 - 76.98)		
0.20	99.39 (99.20 - 99.53)			36.78 (34.76 - 38.84)			36.14 (34.40 - 37.92)			76.22 (75.82 - 76.62)		
0.50	47.07 (43.50 - 50.67)			28.68 (27.07 - 30.34)			28.53 (27.38 - 29.70)			30.29 (25.55 - 35.50)		
1.00	32.74 (30.83 - 34.71)			22.07 (21.22 - 22.95)			23.12 (22.29 - 23.98)			16.60 (15.21 - 18.09)		
2.00	22.36 (21.15 - 23.61)			15.35 (14.65 - 16.08)			17.29 (16.64 - 17.95)			11.16 (10.58 - 11.75)		
5.00	12.63 (12.23 - 13.05)			7.87 (7.52 - 8.24)			8.42 (8.00 - 8.86)			6.18 (5.96 - 6.42)		
10.00	6.36 (6.04 - 6.71)			4.21 (3.96 - 4.48)			3.89 (3.68 - 4.12)			3.37 (3.20 - 3.54)		
20.00	1.62 (1.48 - 1.76)			1.82 (1.70 - 1.94)			1.20 (1.10 - 1.30)			1.08 (0.96 - 1.20)		
50.00	0.11 (0.08 - 0.14)			0.32 (0.28 - 0.37)			0.04 (0.02 - 0.06)			0.11 (0.08 - 0.14)		
Spam Misc%	Ham Misc%											
0.01	100.00 (nan - nan)			86.82 (84.31 - 88.98)			70.66 (62.59 - 77.61)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			84.34 (81.19 - 87.06)			61.80 (50.76 - 71.74)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			80.28 (77.76 - 82.58)			46.66 (43.30 - 50.06)			100.00 (nan - nan)		
0.10	100.00 (nan - nan)			72.47 (68.99 - 75.69)			41.19 (37.86 - 44.60)			100.00 (nan - nan)		
0.20	32.24 (31.36 - 33.14)			61.11 (58.45 - 63.71)			33.97 (32.16 - 35.83)			30.40 (28.55 - 32.32)		
0.50	26.88 (26.12 - 27.65)			40.58 (38.24 - 42.98)			26.72 (25.84 - 27.62)			24.57 (23.83 - 25.33)		
1.00	23.10 (22.43 - 23.78)			27.92 (26.74 - 29.14)			21.44 (20.67 - 22.24)			20.40 (19.74 - 21.09)		
2.00	18.42 (17.73 - 19.13)			18.56 (17.77 - 19.37)			15.62 (15.03 - 16.22)			15.28 (14.83 - 15.74)		
5.00	11.93 (11.48 - 12.40)			8.49 (8.04 - 8.97)			8.29 (7.95 - 8.64)			6.59 (6.25 - 6.95)		
10.00	6.73 (6.42 - 7.05)			3.77 (3.56 - 3.98)			4.19 (3.98 - 4.42)			2.27 (2.11 - 2.44)		
20.00	2.42 (2.24 - 2.62)			1.26 (1.14 - 1.38)			1.45 (1.33 - 1.59)			0.82 (0.73 - 0.91)		
50.00	0.44 (0.38 - 0.52)			0.04 (0.02 - 0.06)			0.03 (0.02 - 0.05)			0.36 (0.30 - 0.43)		

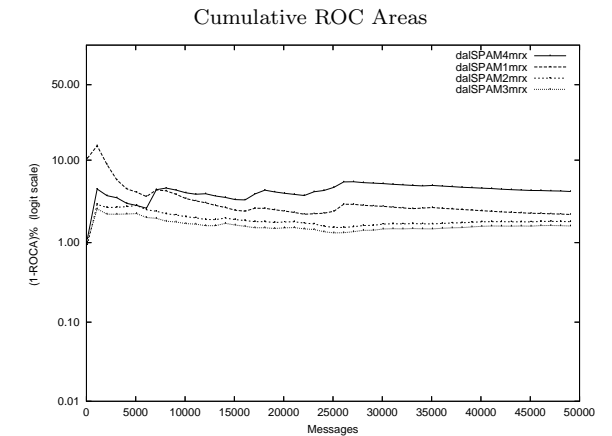
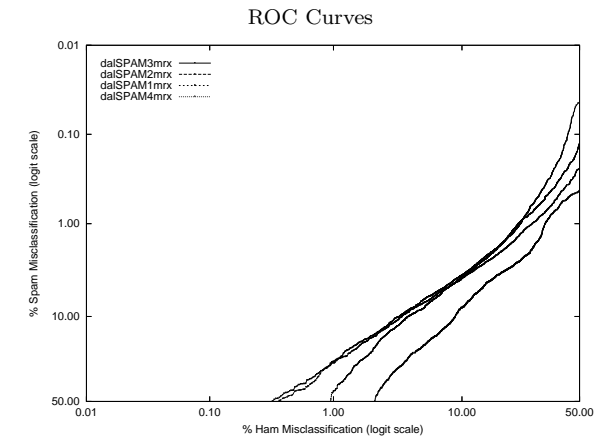


Learning Curves

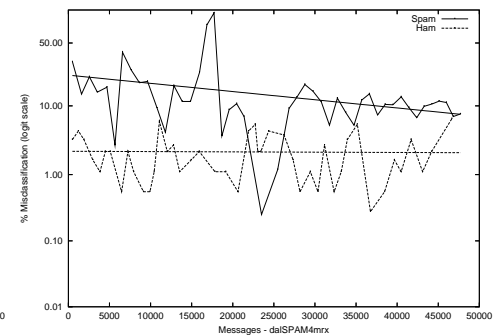
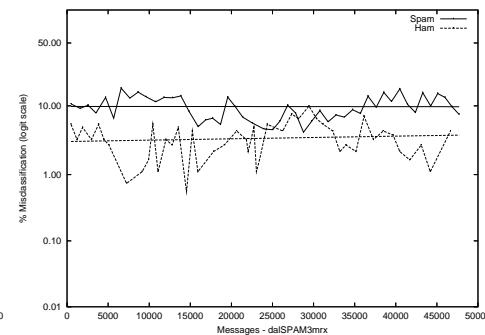
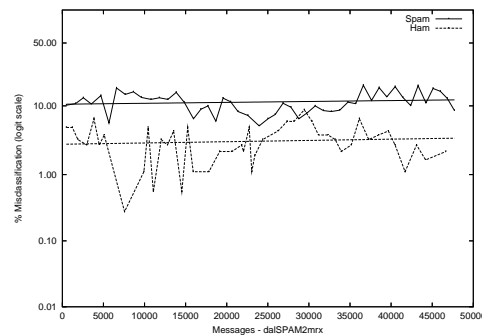
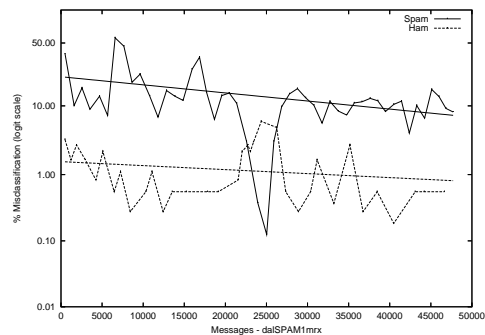


dal - Dalhousie U. on Private Mr X. Corpus

Gold Std. Filter Classification	dalSPAM1mrx		dalSPAM2mrx		dalSPAM3mrx		dalSPAM4mrx	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	1.17 (0.96-1.42)		3.12 (2.77-3.50)		3.44 (3.07-3.84)		2.18 (1.89-2.50)	
Spam Misc%	13.83 (13.50-14.18)		11.33 (11.02-11.65)		9.79 (9.50-10.09)		14.40 (14.06-14.75)	
Logit av Misc%	4.18 (3.81-4.59)		6.03 (5.69-6.39)		5.86 (5.54-6.19)		5.77 (5.39-6.17)	
(1-ROCA)%	2.240 (2.055 - 2.443)		1.824 (1.705 - 1.953)		1.613 (1.518 - 1.714)		4.282 (4.013 - 4.569)	
Ham Misc%	Spam Misc%							
0.01	99.44 (99.34 - 99.52)		99.95 (98.33 - 100.00)		99.97 (97.48 - 100.00)		98.63 (97.93 - 99.10)	
0.02	99.43 (99.32 - 99.51)		98.74 (56.23 - 99.98)		98.21 (40.46 - 99.98)		98.17 (97.21 - 98.81)	
0.05	99.41 (99.28 - 99.51)		85.09 (32.45 - 98.55)		84.54 (33.49 - 98.34)		97.49 (96.66 - 98.12)	
0.10	99.31 (99.18 - 99.43)		69.41 (58.30 - 78.64)		70.03 (48.55 - 85.26)		96.93 (96.01 - 97.65)	
0.20	99.15 (97.62 - 99.70)		57.52 (50.04 - 64.67)		57.14 (48.12 - 65.71)		95.12 (93.65 - 96.26)	
0.50	90.86 (81.99 - 95.59)		43.31 (38.63 - 48.10)		40.69 (34.91 - 46.74)		93.21 (92.14 - 94.13)	
1.00	43.74 (28.83 - 59.88)		27.07 (23.38 - 31.11)		26.35 (22.53 - 30.56)		89.29 (87.74 - 90.66)	
2.00	22.29 (18.48 - 26.64)		16.36 (14.66 - 18.21)		15.64 (14.32 - 17.05)		56.04 (44.02 - 67.38)	
5.00	8.55 (7.87 - 9.28)		7.52 (6.95 - 8.13)		6.99 (6.38 - 7.66)		19.95 (18.37 - 21.62)	
10.00	3.90 (3.61 - 4.22)		4.03 (3.71 - 4.36)		3.72 (3.49 - 3.95)		8.20 (7.54 - 8.90)	
20.00	1.55 (1.38 - 1.73)		1.90 (1.73 - 2.08)		1.54 (1.39 - 1.71)		3.32 (3.09 - 3.55)	
50.00	0.13 (0.09 - 0.18)		0.24 (0.20 - 0.29)		0.04 (0.03 - 0.08)		0.43 (0.37 - 0.50)	
Spam Misc%	Ham Misc%							
0.01	74.66 (60.90 - 84.79)		72.15 (65.09 - 78.26)		59.87 (51.35 - 67.83)		89.36 (88.41 - 90.23)	
0.02	70.58 (64.73 - 75.82)		69.14 (66.43 - 71.73)		53.70 (48.70 - 58.62)		89.13 (88.44 - 89.79)	
0.05	60.57 (53.98 - 66.79)		65.40 (62.34 - 68.34)		47.98 (44.58 - 51.39)		88.79 (87.73 - 89.78)	
0.10	53.36 (49.97 - 56.72)		59.85 (57.00 - 62.63)		44.56 (42.52 - 46.61)		86.40 (84.19 - 88.35)	
0.20	46.81 (43.87 - 49.78)		53.82 (50.98 - 56.63)		40.43 (37.96 - 42.95)		74.87 (66.54 - 81.70)	
0.50	34.70 (32.79 - 36.66)		40.12 (38.28 - 41.99)		31.88 (30.19 - 33.61)		45.38 (41.75 - 49.05)	
1.00	24.24 (22.95 - 25.58)		29.29 (27.29 - 31.37)		25.01 (23.75 - 26.30)		34.60 (33.15 - 36.08)	
2.00	17.07 (16.07 - 18.13)		19.25 (18.02 - 20.54)		16.77 (15.71 - 17.90)		28.76 (27.44 - 30.11)	
5.00	8.02 (7.40 - 8.69)		7.99 (7.37 - 8.65)		7.47 (6.92 - 8.06)		14.18 (13.37 - 15.04)	
10.00	3.99 (3.58 - 4.45)		3.52 (3.14 - 3.94)		3.35 (3.00 - 3.74)		8.97 (8.40 - 9.59)	
20.00	2.15 (1.85 - 2.49)		1.57 (1.34 - 1.85)		1.41 (1.19 - 1.66)		4.98 (4.57 - 5.42)	
50.00	0.95 (0.77 - 1.17)		0.37 (0.27 - 0.50)		0.32 (0.22 - 0.47)		2.14 (1.86 - 2.45)	

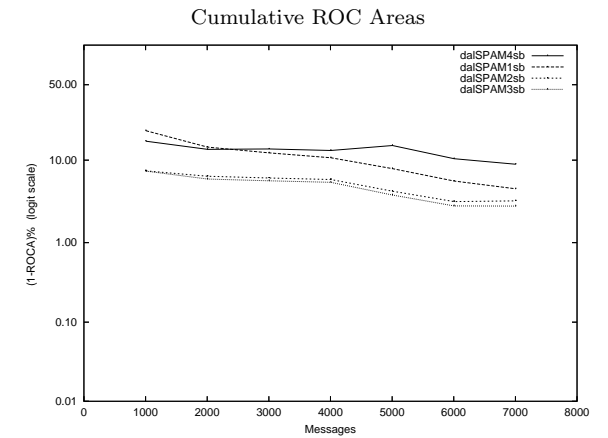
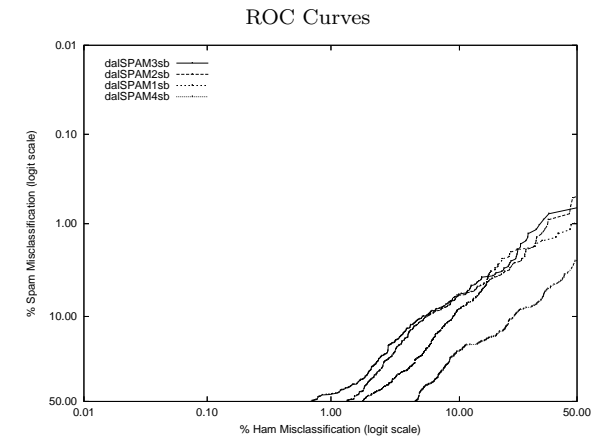


Learning Curves

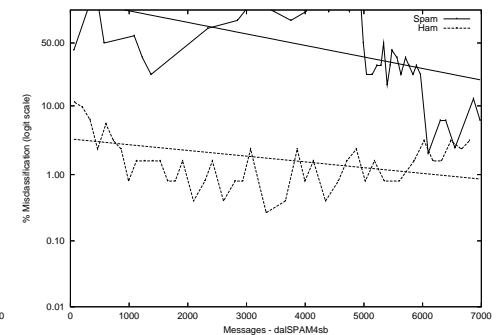
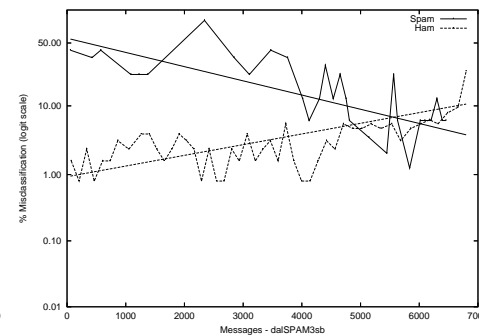
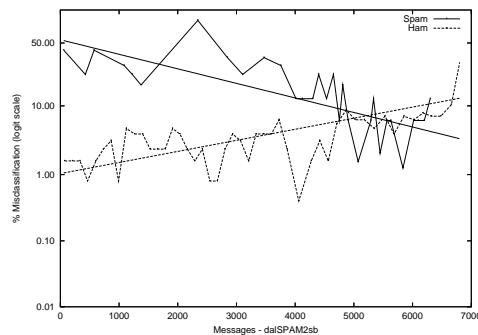
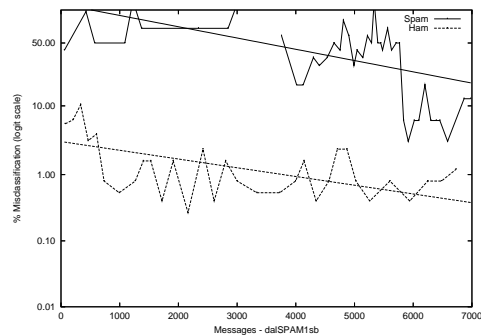


dal - Dalhousie U. on Private S.B. Corpus

Gold Std. Filter Classification	dalSPAM1sb		dalSPAM2sb		dalSPAM3sb		dalSPAM4sb	
	ham	spam	ham	spam	ham	spam	ham	spam
Ham Misc%	1.35	(1.08-1.67)	4.73	(4.22-5.29)	4.03	(3.55-4.55)	1.89	(1.57-2.26)
Spam Misc%	38.19	(34.76-41.72)	12.39	(10.15-14.91)	13.29	(10.98-15.88)	40.90	(37.42-44.46)
Logit av Misc%	8.42	(7.47-9.47)	7.73	(6.91-8.65)	7.43	(6.63-8.30)	10.36	(9.33-11.49)
(1-ROCA)%	4.614	(3.954 - 5.379)	3.293	(2.740 - 3.952)	2.845	(2.405 - 3.364)	9.002	(8.153 - 9.930)
Ham Misc%	Spam Misc%							
0.01	100.00	(nan - nan)	99.48	(89.46 - 99.98)	99.61	(nan - nan)	100.00	(nan - nan)
0.02	100.00	(nan - nan)	90.19	(33.82 - 99.40)	88.00	(nan - nan)	100.00	(nan - nan)
0.05	100.00	(nan - nan)	89.94	(72.18 - 96.85)	83.23	(nan - nan)	100.00	(nan - nan)
0.10	100.00	(nan - nan)	83.48	(75.07 - 89.46)	77.16	(65.72 - 85.62)	100.00	(nan - nan)
0.20	100.00	(nan - nan)	81.03	(75.26 - 85.71)	70.06	(62.26 - 76.86)	100.00	(nan - nan)
0.50	99.61	(nan - nan)	71.87	(65.97 - 77.11)	56.52	(50.57 - 62.28)	87.48	(nan - nan)
1.00	87.23	(40.03 - 98.59)	57.94	(50.32 - 65.19)	45.55	(41.57 - 49.59)	78.45	(74.40 - 82.01)
2.00	46.45	(40.81 - 52.19)	37.94	(30.41 - 46.09)	32.13	(25.80 - 39.19)	66.97	(61.72 - 71.82)
5.00	24.13	(20.65 - 27.99)	11.61	(9.09 - 14.72)	10.71	(8.67 - 13.16)	44.90	(37.68 - 52.35)
10.00	8.52	(6.53 - 11.03)	6.06	(4.48 - 8.17)	5.94	(4.45 - 7.87)	21.42	(18.62 - 24.51)
20.00	2.45	(1.59 - 3.75)	3.61	(2.49 - 5.21)	3.35	(2.32 - 4.82)	13.94	(11.74 - 16.46)
50.00	1.03	(0.52 - 2.05)	0.52	(0.14 - 1.88)	0.77	(0.34 - 1.78)	2.58	(1.67 - 3.96)
Spam Misc%	Ham Misc%							
0.01	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.02	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.05	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.10	98.86	(94.92 - 99.75)	75.64	(65.00 - 83.84)	78.74	(68.27 - 86.44)	95.81	(73.87 - 99.46)
0.20	98.57	(85.45 - 99.88)	70.09	(52.23 - 83.39)	72.99	(56.04 - 85.14)	74.50	(35.15 - 94.03)
0.50	76.02	(18.89 - 97.74)	61.60	(37.57 - 81.04)	58.71	(36.36 - 77.96)	69.39	(62.21 - 75.74)
1.00	51.71	(24.90 - 77.57)	36.88	(22.55 - 53.97)	34.79	(18.78 - 55.19)	65.30	(62.59 - 67.92)
2.00	24.65	(12.87 - 42.01)	27.78	(21.01 - 35.75)	24.88	(19.96 - 30.54)	59.38	(47.43 - 70.31)
5.00	15.21	(13.08 - 17.63)	13.45	(8.92 - 19.79)	12.10	(8.93 - 16.21)	37.84	(32.18 - 43.85)
10.00	9.12	(7.66 - 10.82)	5.84	(4.50 - 7.55)	5.36	(4.09 - 6.99)	23.69	(20.97 - 26.64)
20.00	5.92	(5.22 - 6.71)	3.32	(2.78 - 3.96)	2.86	(2.45 - 3.33)	10.64	(8.34 - 13.48)
50.00	1.80	(1.41 - 2.28)	1.35	(1.01 - 1.80)	0.71	(0.51 - 0.99)	4.64	(4.04 - 5.32)

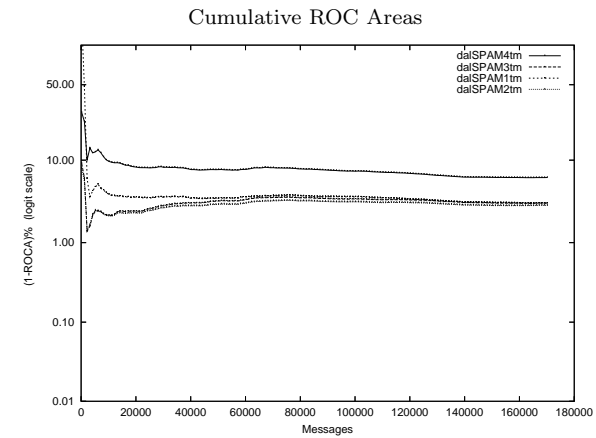
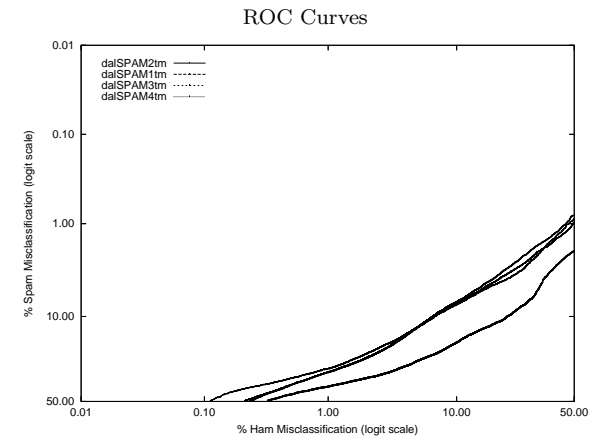


Learning Curves

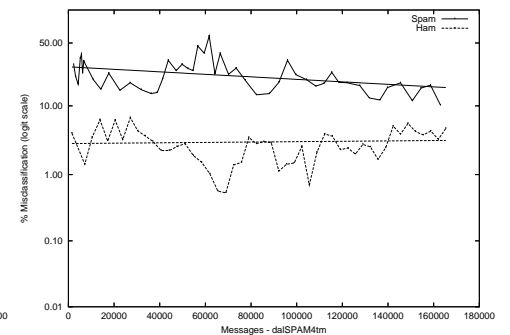
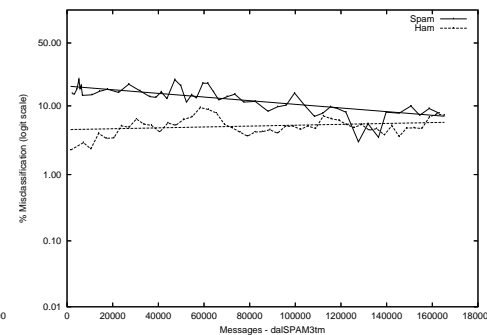
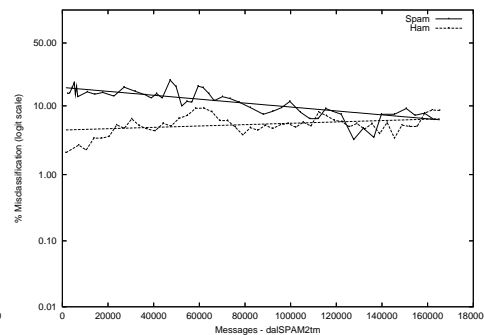
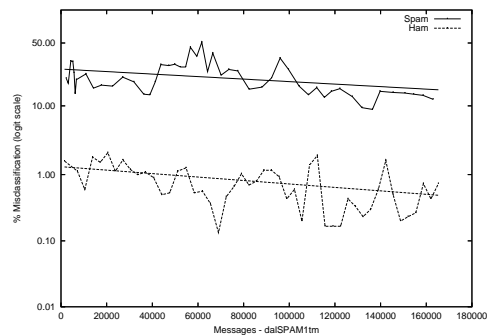


dal - Dalhousie U. on Private T.M. Corpus

Gold Std. Filter Classification	dalSPAM1tm		dalSPAM2tm		dalSPAM3tm		dalSPAM4tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	149454 1231	4454 15062	142281 8404	2308 17208	142746 7939	2465 17051	146066 4619	4728 14788
Ham Misc%	0.82 (0.77-0.86)		5.58 (5.46-5.69)		5.27 (5.16-5.38)		3.07 (2.98-3.15)	
Spam Misc%	22.82 (22.24-23.42)		11.83 (11.38-12.29)		12.63 (12.17-13.10)		24.23 (23.63-24.83)	
Logit av Misc%	4.70 (4.56-4.85)		8.17 (7.99-8.36)		8.23 (8.05-8.41)		9.14 (8.96-9.32)	
(1-ROCA)%	3.085 (2.968 - 3.207)		2.898 (2.792 - 3.007)		3.090 (2.983 - 3.201)		6.294 (6.127 - 6.466)	
Ham Misc%	Spam Misc%							
0.01	99.37 (99.23 - 99.49)		87.61 (82.64 - 91.31)		87.46 (81.05 - 91.92)		66.74 (63.46 - 69.87)	
0.02	99.31 (99.15 - 99.44)		80.68 (74.20 - 85.85)		75.81 (69.92 - 80.87)		66.16 (65.54 - 66.77)	
0.05	64.48 (54.90 - 73.01)		67.49 (65.20 - 69.70)		65.70 (63.39 - 67.94)		65.94 (64.00 - 67.82)	
0.10	52.08 (49.75 - 54.39)		59.84 (58.01 - 61.64)		59.70 (57.49 - 61.88)		58.51 (57.01 - 59.98)	
0.20	42.53 (41.48 - 43.59)		51.85 (50.45 - 53.25)		50.73 (49.01 - 52.44)		53.69 (52.64 - 54.74)	
0.50	35.77 (34.98 - 36.56)		39.75 (38.52 - 40.99)		39.76 (38.72 - 40.81)		45.80 (45.00 - 46.60)	
1.00	30.21 (29.59 - 30.83)		32.16 (31.35 - 32.98)		32.20 (31.50 - 32.91)		40.83 (40.07 - 41.59)	
2.00	22.90 (22.26 - 23.55)		24.72 (24.00 - 25.46)		24.96 (24.35 - 25.59)		35.75 (35.07 - 36.43)	
5.00	13.01 (12.56 - 13.48)		13.07 (12.51 - 13.64)		13.23 (12.74 - 13.73)		26.39 (25.79 - 27.01)	
10.00	7.65 (7.27 - 8.03)		7.15 (6.79 - 7.52)		7.44 (7.13 - 7.76)		18.08 (17.57 - 18.60)	
20.00	4.42 (4.16 - 4.69)		3.59 (3.36 - 3.83)		4.04 (3.80 - 4.29)		11.32 (10.92 - 11.73)	
50.00	0.89 (0.75 - 1.06)		0.81 (0.70 - 0.94)		0.98 (0.87 - 1.11)		2.02 (1.84 - 2.22)	
Spam Misc%	Ham Misc%							
0.01	99.41 (98.74 - 99.73)		94.07 (69.41 - 99.11)		96.89 (86.78 - 99.33)		98.89 (95.65 - 99.72)	
0.02	98.96 (95.36 - 99.78)		92.78 (83.97 - 96.92)		94.33 (87.26 - 97.59)		97.90 (95.38 - 99.06)	
0.05	93.51 (88.77 - 96.33)		83.40 (78.84 - 87.13)		89.17 (84.27 - 92.68)		96.36 (94.86 - 97.44)	
0.10	86.96 (77.54 - 92.79)		78.34 (73.54 - 82.48)		83.69 (79.51 - 87.16)		94.14 (92.19 - 95.62)	
0.20	78.42 (73.77 - 82.45)		70.22 (66.69 - 73.53)		75.36 (71.14 - 79.14)		90.21 (87.09 - 92.64)	
0.50	63.24 (60.14 - 66.23)		57.37 (53.90 - 60.76)		61.84 (58.59 - 64.99)		79.16 (76.24 - 81.82)	
1.00	47.70 (44.33 - 51.08)		46.27 (43.94 - 48.62)		49.56 (47.37 - 51.75)		66.50 (63.39 - 69.47)	
2.00	35.29 (33.82 - 36.79)		30.72 (28.89 - 32.62)		34.70 (32.23 - 37.25)		50.33 (48.08 - 52.59)	
5.00	16.88 (15.77 - 18.06)		14.80 (14.01 - 15.63)		15.92 (15.15 - 16.73)		34.35 (33.77 - 34.94)	
10.00	6.96 (6.68 - 7.25)		6.70 (6.38 - 7.04)		6.71 (6.42 - 7.01)		22.49 (21.64 - 23.37)	
20.00	2.59 (2.45 - 2.72)		2.97 (2.80 - 3.14)		2.97 (2.84 - 3.11)		8.63 (8.31 - 8.97)	
50.00	0.11 (0.10 - 0.13)		0.23 (0.21 - 0.26)		0.21 (0.19 - 0.24)		0.32 (0.29 - 0.36)	



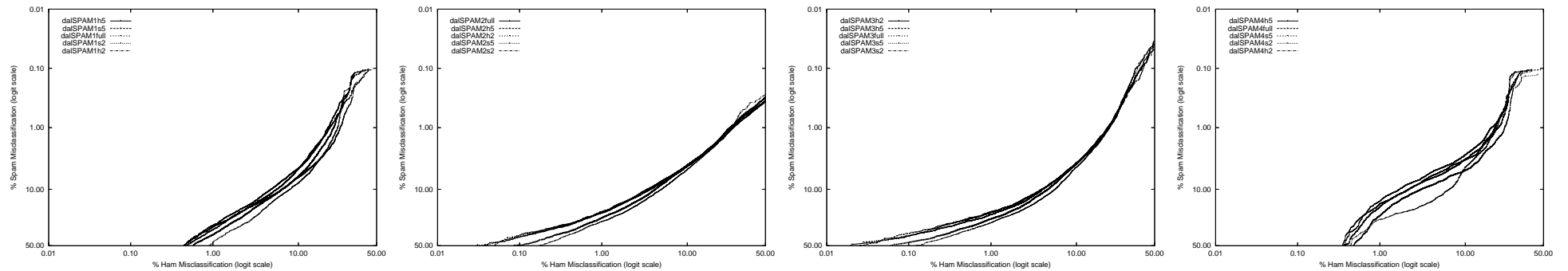
Learning Curves



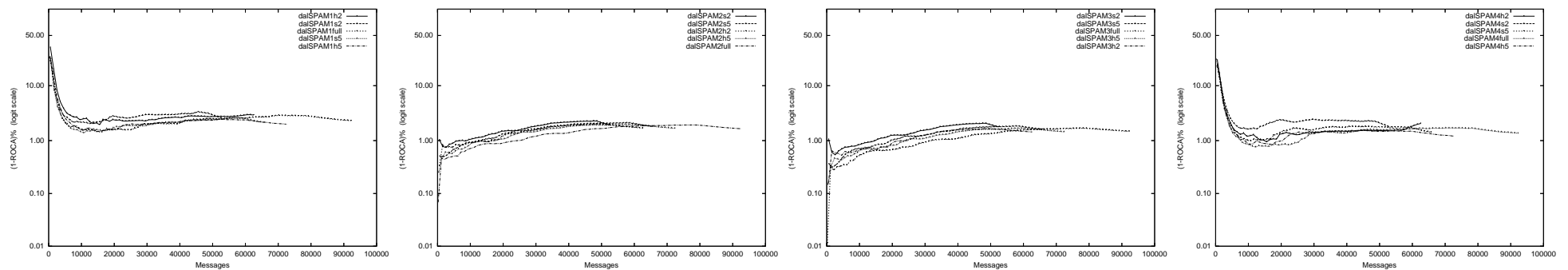
dal - Dalhousie U. - Public subset Corpora

	dalSPAM1full	dalSPAM1h2	dalSPAM1h5	dalSPAM1s2	dalSPAM1s5
Ham Misc%	1.17 (1.07-1.28)	2.27 (1.98-2.58)	1.54 (1.37-1.72)	1.17 (1.07-1.28)	1.09 (0.99-1.20)
Spam Misc%	21.07 (20.73-21.42)	20.67 (20.33-21.02)	17.57 (17.25-17.90)	22.66 (21.95-23.39)	21.26 (20.77-21.76)
Logit av Misc% (1-ROCA)%	5.33 (5.10-5.57)	7.21 (6.77-7.68)	5.46 (5.17-5.77)	5.57 (5.31-5.84)	5.18 (4.94-5.43)
	2.348 (2.255 - 2.444)	2.997 (2.813 - 3.194)	2.026 (1.924 - 2.133)	2.662 (2.540 - 2.790)	2.183 (2.085 - 2.285)
	dalSPAM2full	dalSPAM2h2	dalSPAM2h5	dalSPAM2s2	dalSPAM2s5
Ham Misc%	5.34 (5.12-5.56)	5.88 (5.42-6.36)	5.34 (5.03-5.66)	5.69 (5.47-5.93)	5.65 (5.42-5.88)
Spam Misc%	7.52 (7.29-7.74)	7.11 (6.89-7.33)	7.31 (7.09-7.54)	8.97 (8.49-9.47)	7.88 (7.56-8.21)
Logit av Misc% (1-ROCA)%	6.34 (6.18-6.50)	6.47 (6.20-6.75)	6.25 (6.05-6.46)	7.16 (6.92-7.41)	6.68 (6.49-6.87)
	1.674 (1.609 - 1.742)	1.713 (1.624 - 1.808)	1.694 (1.616 - 1.776)	1.970 (1.854 - 2.093)	1.827 (1.744 - 1.914)
	dalSPAM3full	dalSPAM3h2	dalSPAM3h5	dalSPAM3s2	dalSPAM3s5
Ham Misc%	6.80 (6.56-7.06)	7.11 (6.60-7.64)	7.02 (6.67-7.39)	6.96 (6.71-7.22)	6.94 (6.69-7.19)
Spam Misc%	6.23 (6.03-6.44)	5.88 (5.68-6.09)	6.00 (5.80-6.21)	7.58 (7.13-8.05)	6.48 (6.19-6.79)
Logit av Misc% (1-ROCA)%	6.51 (6.35-6.67)	6.47 (6.22-6.73)	6.49 (6.30-6.69)	7.27 (7.02-7.52)	6.71 (6.51-6.91)
	1.491 (1.441 - 1.544)	1.453 (1.363 - 1.549)	1.459 (1.389 - 1.534)	1.814 (1.729 - 1.904)	1.635 (1.572 - 1.700)
	dalSPAM4full	dalSPAM4h2	dalSPAM4h5	dalSPAM4s2	dalSPAM4s5
Ham Misc%	2.69 (2.54-2.86)	4.66 (4.25-5.09)	3.58 (3.32-3.85)	2.47 (2.32-2.63)	2.28 (2.14-2.44)
Spam Misc%	4.50 (4.33-4.68)	5.44 (5.24-5.63)	3.42 (3.26-3.57)	6.19 (5.79-6.62)	4.88 (4.62-5.15)
Logit av Misc% (1-ROCA)%	3.49 (3.36-3.61)	5.03 (4.80-5.28)	3.50 (3.35-3.65)	3.93 (3.75-4.11)	3.35 (3.21-3.49)
	1.370 (1.306 - 1.436)	2.087 (1.940 - 2.245)	1.217 (1.137 - 1.302)	1.854 (1.736 - 1.979)	1.430 (1.344 - 1.521)

ROC Curves

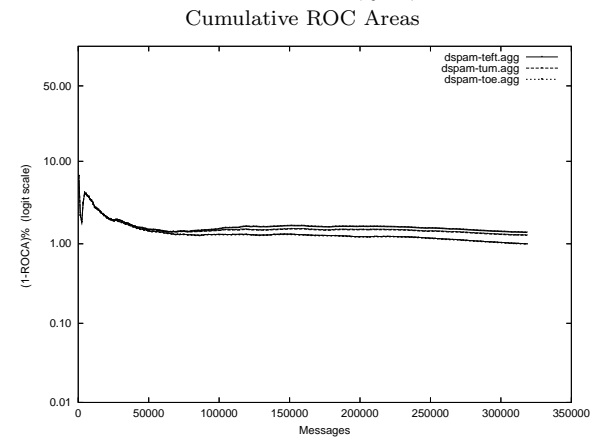
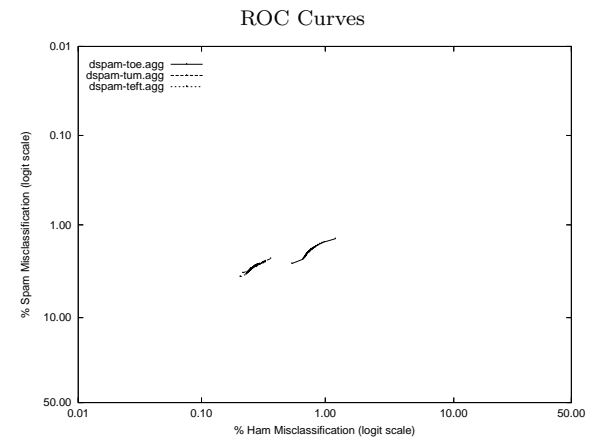


Cumulative ROC Area

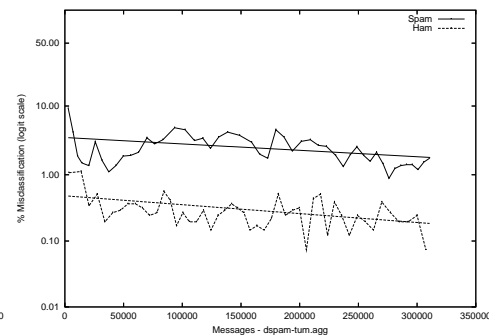
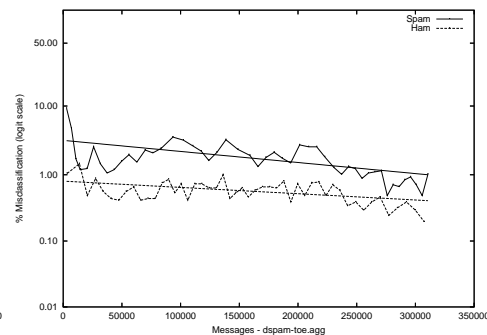
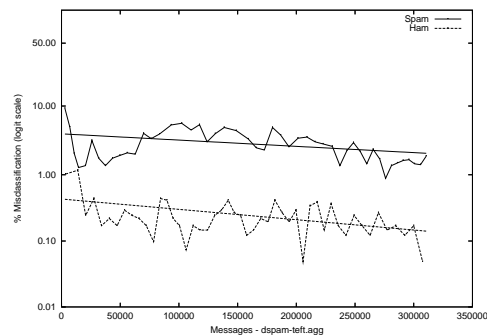


dspam - DSPAM (non-participant) on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	dspam-teft.agg			dspam-toe.agg			dspam-tum.agg		
	ham	spam		ham	spam		ham	spam	
Classification	ham	204825	3316	ham	204163	2130	ham	204722	2907
	spam	528	109813	spam	1190	110999	spam	631	110222
Ham Misc%	0.26 (0.24-0.28)			0.58 (0.55-0.61)			0.31 (0.28-0.33)		
Spam Misc%	2.93 (2.83-3.03)			1.88 (1.80-1.96)			2.57 (2.48-2.66)		
Logit av Misc%	0.87 (0.84-0.92)			1.05 (1.01-1.08)			0.89 (0.86-0.93)		
(1-ROCA)%	1.383 (1.332 - 1.436)			0.987 (0.941 - 1.036)			1.274 (1.225 - 1.325)		
Ham Misc%	Spam Misc%								
0.01	94.20 (91.70 - 95.98)			98.08 (96.83 - 98.84)			93.80 (90.61 - 95.96)		
0.02	90.77 (87.78 - 93.09)			96.31 (94.96 - 97.31)			90.03 (86.32 - 92.82)		
0.05	77.57 (72.86 - 81.66)			92.19 (90.50 - 93.61)			76.64 (71.75 - 80.91)		
0.10	51.60 (45.29 - 57.86)			83.68 (81.22 - 85.87)			51.43 (45.20 - 57.62)		
0.20	6.60 (2.65 - 15.46)			65.05 (61.61 - 68.33)			8.58 (3.65 - 18.87)		
0.50	2.57 (2.49 - 2.66)			9.19 (4.70 - 17.19)			2.35 (2.26 - 2.44)		
1.00	2.56 (2.47 - 2.64)			1.54 (1.46 - 1.61)			2.34 (2.25 - 2.42)		
2.00	2.54 (2.45 - 2.62)			1.42 (1.35 - 1.49)			2.32 (2.23 - 2.41)		
5.00	2.45 (2.37 - 2.54)			1.38 (1.31 - 1.45)			2.24 (2.16 - 2.33)		
10.00	2.32 (2.23 - 2.40)			1.29 (1.22 - 1.35)			2.11 (2.03 - 2.19)		
20.00	2.06 (1.98 - 2.14)			1.14 (1.08 - 1.20)			1.87 (1.80 - 1.95)		
50.00	1.26 (1.20 - 1.33)			0.71 (0.66 - 0.76)			1.16 (1.10 - 1.22)		
Spam Misc%	Ham Misc%								
0.01	99.47 (99.07 - 99.70)			99.00 (98.23 - 99.44)			99.64 (99.39 - 99.79)		
0.02	99.20 (98.76 - 99.48)			98.39 (97.59 - 98.92)			99.23 (98.87 - 99.47)		
0.05	97.90 (97.28 - 98.38)			95.74 (94.35 - 96.80)			97.87 (97.18 - 98.40)		
0.10	96.06 (95.27 - 96.71)			92.28 (90.62 - 93.67)			95.78 (95.00 - 96.44)		
0.20	92.11 (90.96 - 93.13)			86.30 (84.38 - 88.02)			91.21 (90.02 - 92.27)		
0.50	81.34 (79.71 - 82.86)			64.85 (61.78 - 67.79)			79.50 (77.76 - 81.14)		
1.00	60.02 (57.91 - 62.09)			30.92 (26.79 - 35.37)			56.12 (53.62 - 58.59)		
2.00	22.36 (19.45 - 25.56)			0.73 (0.69 - 0.77)			14.65 (11.82 - 18.01)		
5.00	0.20 (0.18 - 0.22)			0.53 (0.50 - 0.56)			0.21 (0.19 - 0.23)		
10.00	0.19 (0.17 - 0.21)			0.50 (0.47 - 0.53)			0.20 (0.18 - 0.22)		
20.00	0.17 (0.15 - 0.19)			0.44 (0.42 - 0.47)			0.17 (0.15 - 0.19)		
50.00	0.11 (0.09 - 0.12)			0.28 (0.26 - 0.30)			0.11 (0.09 - 0.12)		

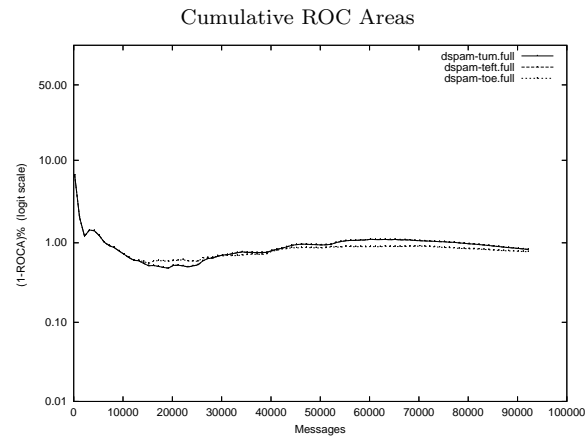
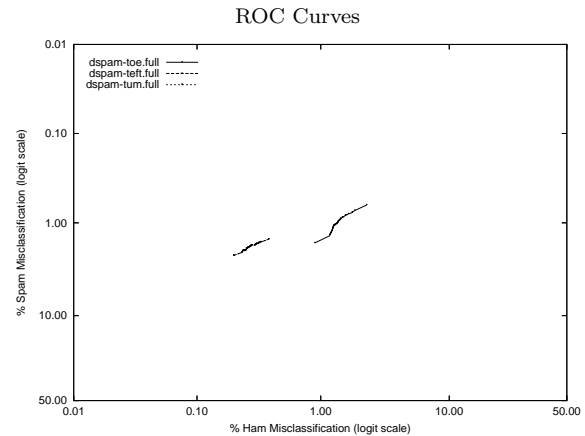


Learning Curves

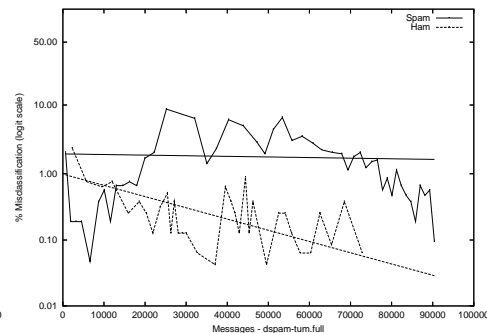
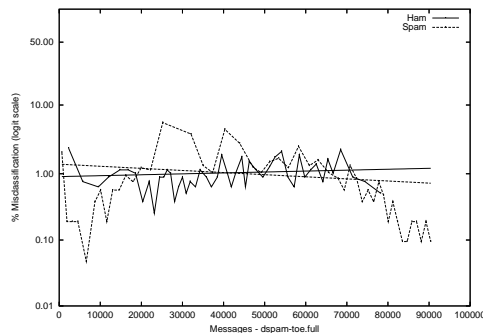
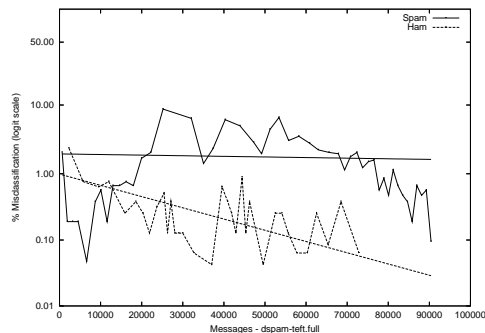


dspam - DSPAM (non-participant) on Public trec05p-1/full Corpus

Gold Std. Filter Classification	dspam-teft.full			dspam-toe.full			dspam-tum.full		
	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	spam
Ham Misc%	0.26 (0.21-0.32)			1.04 (0.94-1.14)			0.26 (0.21-0.32)		
Spam Misc%	1.79 (1.68-1.90)			0.99 (0.91-1.08)			1.79 (1.68-1.90)		
Logit av Misc%	0.69 (0.62-0.76)			1.01 (0.95-1.08)			0.69 (0.62-0.76)		
(1-ROCA)%	0.827 (0.773 - 0.884)			0.773 (0.707 - 0.846)			0.827 (0.773 - 0.884)		
Ham Misc%	Spam Misc%								
0.01	94.07 (83.92 - 97.97)			98.47 (96.18 - 99.39)			94.07 (83.92 - 97.97)		
0.02	88.86 (77.04 - 94.99)			97.41 (94.79 - 98.73)			88.86 (77.04 - 94.99)		
0.05	74.63 (63.21 - 83.43)			93.61 (90.16 - 95.91)			74.63 (63.21 - 83.43)		
0.10	47.09 (33.82 - 60.78)			88.76 (84.24 - 92.10)			47.09 (33.82 - 60.78)		
0.20	2.28 (0.51 - 9.63)			80.56 (75.38 - 84.87)			2.28 (0.51 - 9.63)		
0.50	1.51 (1.42 - 1.61)			44.39 (36.55 - 52.52)			1.51 (1.42 - 1.61)		
1.00	1.51 (1.42 - 1.60)			1.55 (1.21 - 1.98)			1.51 (1.42 - 1.60)		
2.00	1.49 (1.40 - 1.59)			0.70 (0.63 - 0.79)			1.49 (1.40 - 1.59)		
5.00	1.44 (1.35 - 1.53)			0.62 (0.55 - 0.69)			1.44 (1.35 - 1.53)		
10.00	1.34 (1.26 - 1.43)			0.59 (0.53 - 0.66)			1.34 (1.26 - 1.43)		
20.00	1.20 (1.11 - 1.29)			0.53 (0.47 - 0.59)			1.20 (1.11 - 1.29)		
50.00	0.72 (0.66 - 0.79)			0.32 (0.27 - 0.38)			0.72 (0.66 - 0.79)		
Spam Misc%	Ham Misc%								
0.01	99.36 (98.37 - 99.75)			97.12 (94.51 - 98.51)			99.36 (98.37 - 99.75)		
0.02	99.07 (98.32 - 99.49)			94.87 (91.58 - 96.92)			99.07 (98.32 - 99.49)		
0.05	96.19 (94.65 - 97.29)			88.11 (82.78 - 91.96)			96.19 (94.65 - 97.29)		
0.10	92.16 (89.84 - 93.99)			81.72 (76.80 - 85.79)			92.16 (89.84 - 93.99)		
0.20	84.75 (81.70 - 87.36)			69.09 (63.35 - 74.30)			84.75 (81.70 - 87.36)		
0.50	65.83 (62.00 - 69.45)			26.21 (16.64 - 38.73)			65.83 (62.00 - 69.45)		
1.00	32.08 (27.00 - 37.62)			1.37 (1.24 - 1.51)			32.08 (27.00 - 37.62)		
2.00	0.25 (0.21 - 0.30)			0.89 (0.80 - 0.99)			0.25 (0.21 - 0.30)		
5.00	0.20 (0.16 - 0.24)			0.86 (0.77 - 0.96)			0.20 (0.16 - 0.24)		
10.00	0.18 (0.15 - 0.22)			0.80 (0.72 - 0.90)			0.18 (0.15 - 0.22)		
20.00	0.16 (0.13 - 0.19)			0.73 (0.65 - 0.82)			0.16 (0.13 - 0.19)		
50.00	0.09 (0.07 - 0.12)			0.42 (0.36 - 0.50)			0.09 (0.07 - 0.12)		

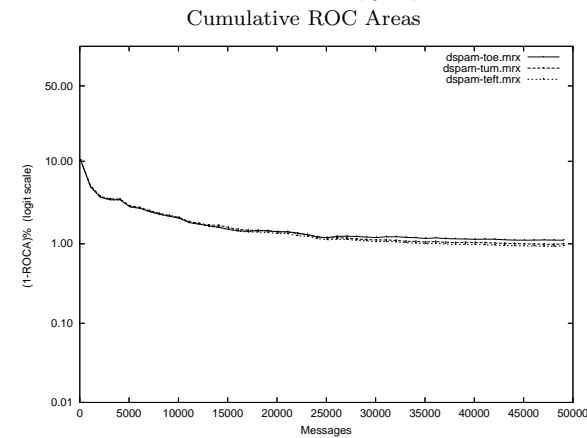
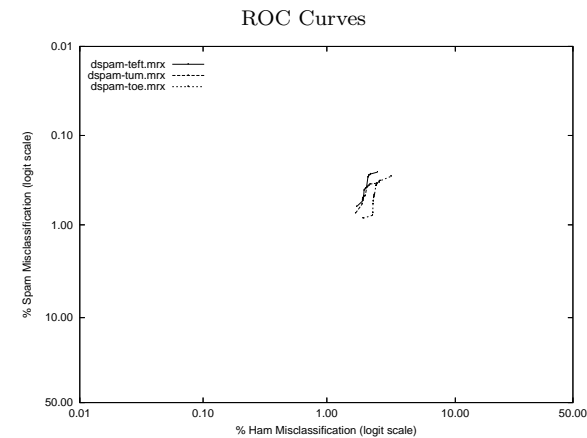


Learning Curves

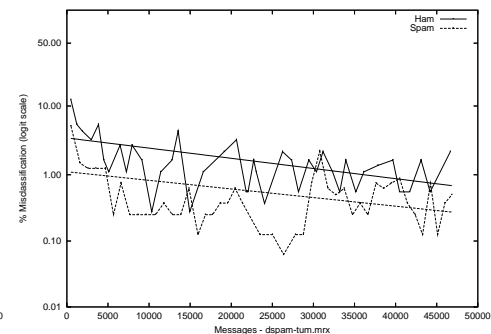
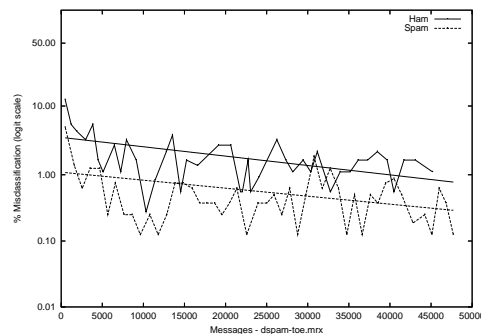
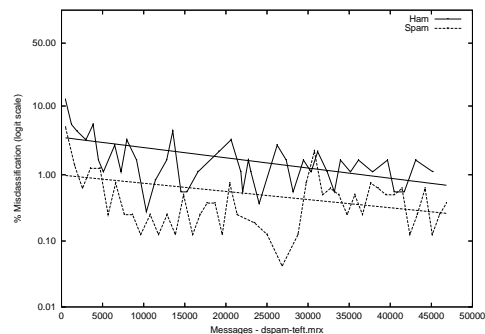


dspam - DSPAM (non-participant) on Private Mr X. Corpus

Gold Std. Filter Classification	dspam-teft.mrx		dspam-toe.mrx		dspam-tum.mrx	
	ham	spam	ham	spam	ham	spam
ham	8871	211	8863	236	8874	230
spam	167	39837	175	39812	164	39818
Ham Misc%	1.85 (1.58-2.15)		1.94 (1.66-2.24)		1.81 (1.55-2.11)	
Spam Misc%	0.53 (0.46-0.60)		0.59 (0.52-0.67)		0.57 (0.50-0.65)	
Logit av Misc%	0.99 (0.89-1.09)		1.07 (0.97-1.18)		1.02 (0.93-1.13)	
(1-ROCA)%	0.942 (0.798 - 1.111)		1.109 (0.944 - 1.303)		0.997 (0.857 - 1.159)	
Ham Misc%	Spam Misc%					
0.01	99.55 (nan - nan)		99.48 (93.47 - 99.96)		99.55 (nan - nan)	
0.02	99.48 (97.54 - 99.89)		99.15 (93.61 - 99.89)		99.48 (96.59 - 99.92)	
0.05	98.36 (95.64 - 99.39)		98.36 (96.12 - 99.32)		98.37 (95.34 - 99.44)	
0.10	95.17 (90.97 - 97.47)		96.23 (93.53 - 97.83)		95.18 (89.53 - 97.85)	
0.20	86.41 (79.76 - 91.12)		89.13 (83.50 - 93.00)		86.77 (78.69 - 92.10)	
0.50	69.91 (60.51 - 77.89)		72.89 (63.98 - 80.28)		70.65 (61.68 - 78.26)	
1.00	35.65 (24.56 - 48.54)		47.28 (36.59 - 58.23)		36.07 (25.68 - 47.94)	
2.00	0.41 (0.14 - 1.17)		0.83 (0.10 - 6.40)		0.48 (0.25 - 0.91)	
5.00	0.26 (0.22 - 0.31)		0.28 (0.24 - 0.34)		0.34 (0.29 - 0.40)	
10.00	0.24 (0.20 - 0.29)		0.27 (0.23 - 0.33)		0.31 (0.26 - 0.37)	
20.00	0.23 (0.19 - 0.28)		0.24 (0.20 - 0.30)		0.30 (0.25 - 0.36)	
50.00	0.15 (0.12 - 0.19)		0.17 (0.13 - 0.21)		0.19 (0.15 - 0.23)	
Spam Misc%	Ham Misc%					
0.01	93.22 (82.11 - 97.63)		92.39 (84.02 - 96.55)		94.82 (88.60 - 97.74)	
0.02	92.17 (85.91 - 95.78)		85.70 (76.40 - 91.74)		93.23 (88.32 - 96.16)	
0.05	80.00 (71.86 - 86.23)		78.07 (68.72 - 85.23)		82.62 (75.33 - 88.09)	
0.10	60.48 (50.80 - 69.40)		64.59 (54.23 - 73.75)		70.06 (61.32 - 77.55)	
0.20	34.21 (16.98 - 56.94)		38.91 (24.77 - 55.21)		45.12 (33.10 - 57.73)	
0.50	1.97 (1.70 - 2.28)		2.37 (2.03 - 2.75)		1.95 (1.69 - 2.25)	
1.00	1.73 (1.47 - 2.02)		1.95 (1.66 - 2.29)		1.69 (1.47 - 1.95)	
2.00	1.71 (1.46 - 2.01)		1.93 (1.64 - 2.26)		1.68 (1.46 - 1.94)	
5.00	1.67 (1.42 - 1.96)		1.86 (1.58 - 2.19)		1.64 (1.41 - 1.90)	
10.00	1.55 (1.32 - 1.82)		1.76 (1.50 - 2.07)		1.53 (1.31 - 1.78)	
20.00	1.33 (1.11 - 1.58)		1.54 (1.30 - 1.82)		1.34 (1.12 - 1.60)	
50.00	0.75 (0.58 - 0.97)		0.94 (0.76 - 1.17)		0.77 (0.62 - 0.97)	

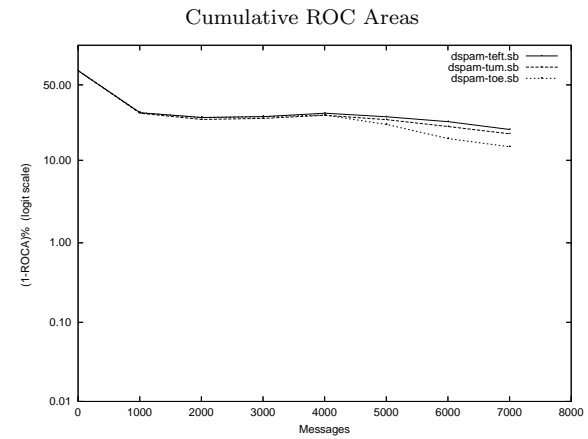
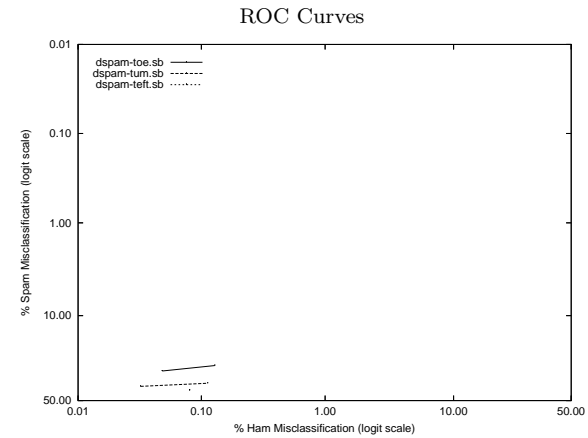


Learning Curves

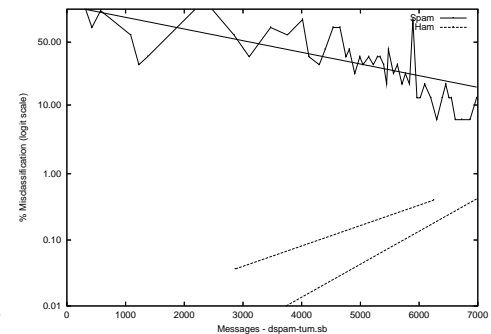
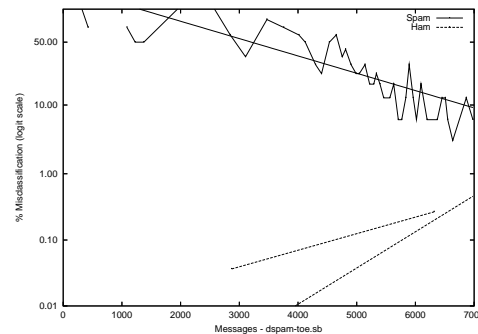
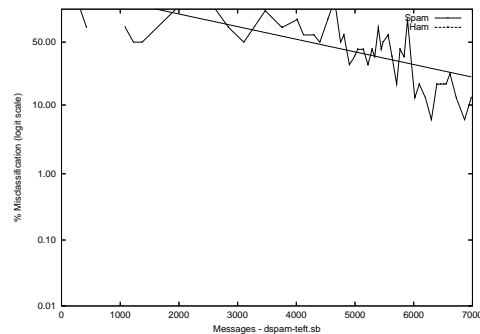


dspam - DSPAM (non-participant) on Private S.B. Corpus

Gold Std. Filter Classification	dspam-teft.sb		dspam-toe.sb		dspam-tum.sb	
	ham	spam	ham	spam	ham	spam
ham	6231	343	6228	240	6228	273
spam	0	432	3	535	3	502
Ham Misc%	0.00 (0.00-0.05)		0.05 (0.01-0.14)		0.05 (0.01-0.14)	
Spam Misc%	44.26 (40.72-47.84)		30.97 (27.73-34.36)		35.23 (31.86-38.70)	
Logit av Misc%	0.00 (0.00-nan)		1.45 (0.82-2.54)		1.59 (0.91-2.78)	
(1-ROCA)%	21.428 (19.356 - 23.656)		14.149 (12.188 - 16.366)		19.384 (17.239 - 21.725)	
Ham Misc%	Spam Misc%					
0.01	45.94 (41.96 - 49.96)		96.65 (nan - nan)		97.68 (nan - nan)	
0.02	45.94 (42.03 - 49.89)		57.03 (9.34 - 94.47)		88.65 (nan - nan)	
0.05	45.81 (41.28 - 50.41)		31.74 (11.39 - 62.72)		40.90 (17.69 - 69.04)	
0.10	43.35 (38.96 - 47.85)		31.61 (23.75 - 40.69)		40.77 (36.91 - 44.75)	
0.20	43.35 (39.46 - 47.34)		28.77 (25.09 - 32.77)		38.97 (35.25 - 42.82)	
0.50	42.97 (39.11 - 46.91)		28.65 (25.45 - 32.07)		38.71 (35.03 - 42.53)	
1.00	42.32 (38.54 - 46.19)		27.87 (24.73 - 31.24)		37.94 (34.20 - 41.82)	
2.00	41.68 (37.91 - 45.55)		27.74 (24.57 - 31.16)		37.29 (33.60 - 41.13)	
5.00	40.26 (36.62 - 44.01)		26.71 (23.63 - 30.04)		35.87 (32.25 - 39.67)	
10.00	38.32 (34.83 - 41.94)		25.55 (22.46 - 28.90)		34.45 (30.77 - 38.33)	
20.00	33.68 (30.37 - 37.16)		21.94 (19.00 - 25.18)		30.32 (27.11 - 33.74)	
50.00	21.29 (18.56 - 24.30)		13.68 (11.29 - 16.47)		19.10 (16.47 - 22.03)	
Spam Misc%	Ham Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.05	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.10	99.87 (nan - nan)		99.87 (nan - nan)		99.87 (nan - nan)	
0.20	99.84 (nan - nan)		99.20 (96.34 - 99.83)		99.20 (96.26 - 99.83)	
0.50	99.15 (97.20 - 99.74)		98.65 (96.15 - 99.54)		98.65 (96.38 - 99.50)	
1.00	98.33 (96.30 - 99.26)		97.77 (95.12 - 99.00)		97.93 (96.06 - 98.92)	
2.00	96.97 (94.75 - 98.27)		95.22 (92.18 - 97.11)		96.42 (94.17 - 97.82)	
5.00	91.03 (87.70 - 93.52)		86.52 (81.36 - 90.42)		89.73 (85.46 - 92.85)	
10.00	78.56 (73.40 - 82.95)		66.91 (58.66 - 74.23)		77.66 (71.55 - 82.77)	
20.00	53.36 (46.63 - 59.98)		24.68 (15.89 - 36.24)		46.67 (38.88 - 54.62)	
50.00	0.01 (0.00 - 0.02)		0.05 (0.01 - 0.20)		0.03 (0.01 - 0.11)	

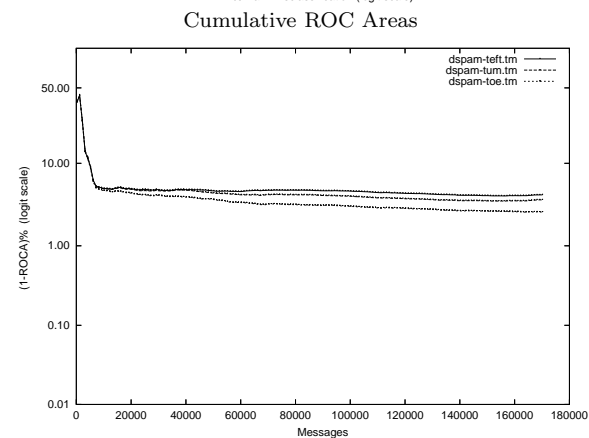
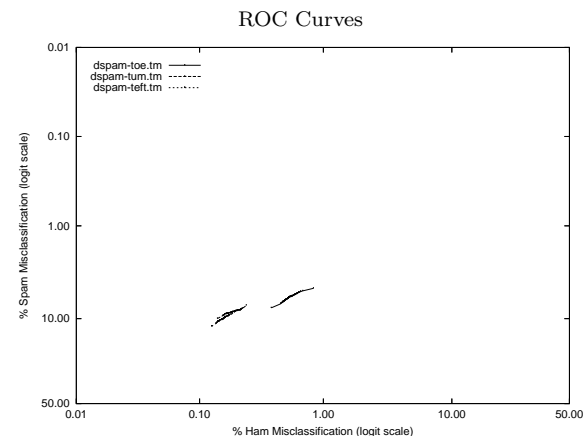


Learning Curves

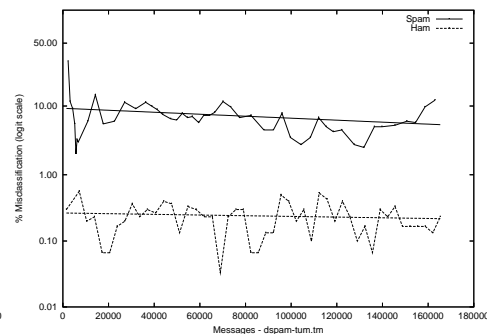
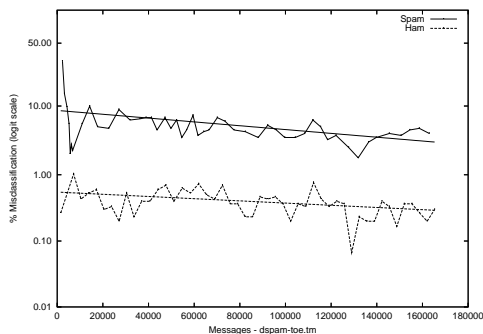
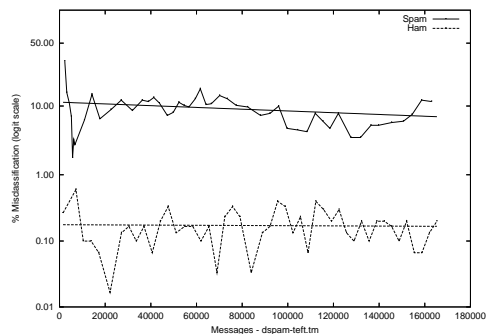


dspam - DSPAM (non-participant) on Private T.M. Corpus

Gold Std. Filter Classification	dspam-teft.tm			dspam-toe.tm			dspam-tum.tm		
	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	
Ham Misc%	0.17 (0.15-0.19)			0.40 (0.37-0.43)			0.24 (0.22-0.27)		
Spam Misc%	9.32 (8.91-9.73)			5.78 (5.46-6.12)			7.48 (7.12-7.86)		
Logit av Misc%	1.31 (1.23-1.40)			1.55 (1.47-1.63)			1.37 (1.30-1.46)		
(1-ROCA)%	4.263 (4.027 - 4.511)			2.626 (2.448 - 2.817)			3.700 (3.499 - 3.912)		
Ham Misc%	Spam Misc%								
0.01	90.54 (85.27 - 94.05)			97.08 (95.47 - 98.13)			94.08 (90.04 - 96.54)		
0.02	85.03 (79.17 - 89.45)			93.51 (91.32 - 95.18)			87.62 (81.84 - 91.74)		
0.05	66.14 (57.39 - 73.91)			85.20 (82.18 - 87.79)			70.52 (62.25 - 77.63)		
0.10	33.79 (22.90 - 46.73)			77.16 (74.11 - 79.94)			37.22 (28.10 - 47.35)		
0.20	8.60 (8.08 - 9.16)			53.62 (48.67 - 58.49)			8.31 (7.78 - 8.87)		
0.50	8.58 (8.13 - 9.06)			6.37 (5.87 - 6.91)			7.29 (6.95 - 7.65)		
1.00	8.55 (8.09 - 9.02)			4.82 (4.52 - 5.14)			7.26 (6.92 - 7.62)		
2.00	8.45 (8.01 - 8.92)			4.79 (4.49 - 5.11)			7.19 (6.85 - 7.55)		
5.00	8.14 (7.70 - 8.59)			4.65 (4.36 - 4.97)			6.94 (6.62 - 7.28)		
10.00	7.72 (7.30 - 8.15)			4.44 (4.14 - 4.75)			6.58 (6.27 - 6.91)		
20.00	6.76 (6.39 - 7.16)			3.90 (3.63 - 4.19)			5.74 (5.45 - 6.04)		
50.00	4.21 (3.92 - 4.52)			2.42 (2.21 - 2.64)			3.66 (3.41 - 3.93)		
Spam Misc%	Ham Misc%								
0.01	99.94 (99.74 - 99.99)			100.00 (99.99 - 100.00)			99.94 (99.68 - 99.99)		
0.02	99.82 (99.48 - 99.93)			99.82 (99.51 - 99.93)			99.82 (99.47 - 99.94)		
0.05	99.55 (99.12 - 99.77)			98.93 (97.99 - 99.43)			99.60 (99.24 - 99.79)		
0.10	98.76 (98.06 - 99.21)			96.97 (95.36 - 98.04)			98.84 (98.13 - 99.28)		
0.20	96.88 (95.84 - 97.67)			94.87 (93.09 - 96.20)			96.88 (95.67 - 97.75)		
0.50	93.23 (91.79 - 94.43)			89.84 (87.60 - 91.70)			93.20 (91.49 - 94.58)		
1.00	86.74 (84.68 - 88.56)			79.52 (76.54 - 82.21)			85.80 (83.59 - 87.76)		
2.00	74.92 (72.51 - 77.17)			58.37 (54.25 - 62.37)			72.86 (70.24 - 75.32)		
5.00	41.01 (37.20 - 44.93)			0.77 (0.33 - 1.75)			30.42 (26.54 - 34.60)		
10.00	0.16 (0.13 - 0.18)			0.37 (0.35 - 0.40)			0.14 (0.12 - 0.16)		
20.00	0.12 (0.10 - 0.14)			0.34 (0.31 - 0.37)			0.13 (0.11 - 0.15)		
50.00	0.07 (0.06 - 0.09)			0.22 (0.20 - 0.24)			0.08 (0.07 - 0.09)		

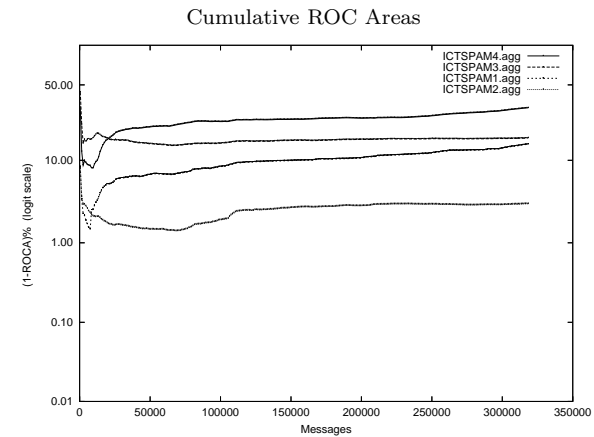
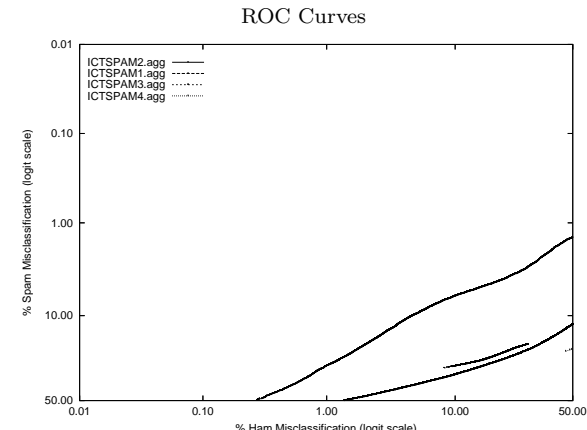


Learning Curves

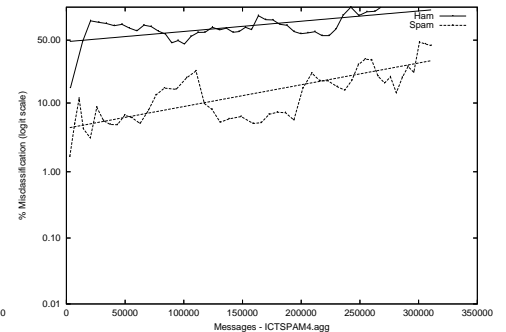
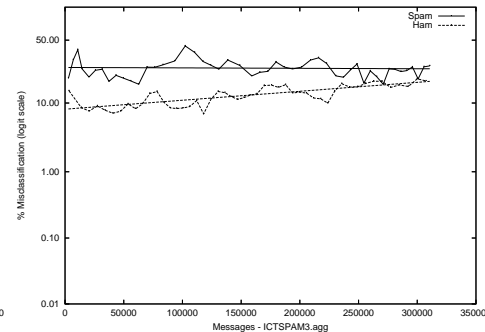
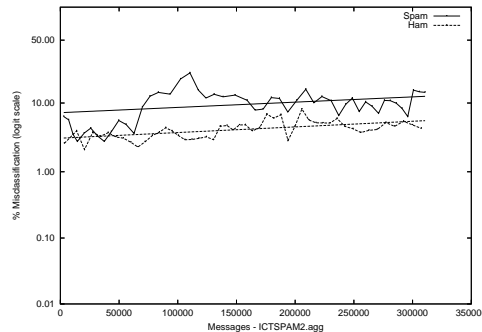
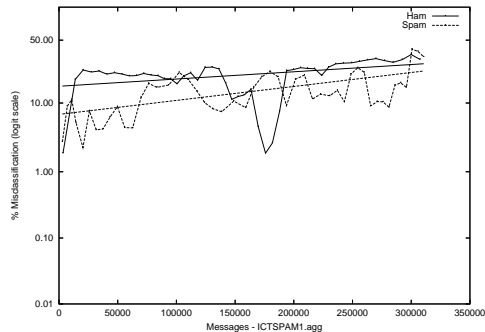


ICT - Chinese Academy of Sciences (ICT) - Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	ICTSPAM1.agg		ICTSPAM2.agg		ICTSPAM3.agg		ICTSPAM4.agg	
	ham	spam	ham	spam	ham	spam	ham	spam
Classification	157796 47557	17195 95934	196499 8854	11082 102047	178436 26917	30916 82213	77749 127604	18126 95003
Ham Misc%	23.16 (22.98-23.34)		4.31 (4.22-4.40)		13.11 (12.96-13.25)		62.14 (61.93-62.35)	
Spam Misc%	15.20 (14.99-15.41)		9.80 (9.62-9.97)		27.33 (27.07-27.59)		16.02 (15.81-16.24)	
Logit av Misc%	18.86 (18.71-19.01)		6.54 (6.45-6.63)		19.24 (19.09-19.38)		35.88 (35.67-36.09)	
(1-ROCA)%	15.115 (14.945 - 15.286)		3.048 (2.988 - 3.110)		17.637 (17.457 - 17.818)		33.879 (33.681 - 34.077)	
Ham Misc%	Spam Misc%							
0.01	90.68 (83.61 - 94.88)		89.69 (77.12 - 95.73)		99.93 (99.89 - 99.95)		99.98 (99.96 - 99.99)	
0.02	81.01 (78.16 - 83.56)		79.45 (76.89 - 81.78)		99.86 (99.80 - 99.90)		99.97 (99.95 - 99.98)	
0.05	74.13 (72.38 - 75.81)		69.56 (67.35 - 71.68)		99.61 (99.51 - 99.69)		99.93 (99.91 - 99.95)	
0.10	67.60 (66.61 - 68.57)		60.29 (58.58 - 61.98)		99.17 (99.03 - 99.29)		99.84 (99.80 - 99.87)	
0.20	62.90 (62.12 - 63.67)		52.50 (51.81 - 53.18)		98.26 (98.06 - 98.45)		99.69 (99.64 - 99.73)	
0.50	56.87 (56.37 - 57.38)		41.00 (39.99 - 42.02)		95.77 (95.50 - 96.03)		99.18 (99.11 - 99.25)	
1.00	52.20 (51.76 - 52.63)		28.87 (28.24 - 29.50)		91.61 (91.22 - 91.99)		98.34 (98.23 - 98.45)	
2.00	47.41 (47.02 - 47.79)		19.34 (18.95 - 19.73)		83.31 (82.79 - 83.82)		96.69 (96.55 - 96.82)	
5.00	40.03 (39.69 - 40.37)		9.72 (9.51 - 9.94)		58.04 (57.15 - 58.93)		91.64 (91.42 - 91.84)	
10.00	33.79 (33.44 - 34.14)		6.26 (6.12 - 6.40)		29.04 (28.75 - 29.33)		83.15 (82.87 - 83.43)	
20.00	26.44 (26.12 - 26.77)		4.36 (4.24 - 4.48)		23.40 (23.14 - 23.66)		66.65 (66.24 - 67.06)	
50.00	12.28 (12.08 - 12.49)		1.46 (1.39 - 1.53)		14.17 (13.97 - 14.38)		20.87 (20.60 - 21.14)	
Spam Misc%	Ham Misc%							
0.01	100.00 (nan - nan)		97.90 (96.59 - 98.72)		99.96 (99.93 - 99.98)		99.96 (99.93 - 99.98)	
0.02	100.00 (nan - nan)		96.38 (94.08 - 97.80)		99.94 (99.90 - 99.96)		99.93 (99.89 - 99.96)	
0.05	100.00 (nan - nan)		91.43 (90.51 - 92.27)		99.82 (99.77 - 99.86)		99.82 (99.76 - 99.87)	
0.10	99.78 (99.40 - 99.92)		88.73 (88.18 - 89.26)		99.61 (99.54 - 99.68)		99.67 (99.60 - 99.73)	
0.20	98.35 (97.65 - 98.85)		84.41 (83.22 - 85.52)		99.30 (99.20 - 99.38)		99.37 (99.26 - 99.46)	
0.50	89.37 (88.30 - 90.36)		73.86 (72.90 - 74.78)		98.31 (98.17 - 98.44)		98.36 (98.18 - 98.51)	
1.00	84.16 (83.83 - 84.50)		62.23 (60.83 - 63.60)		96.50 (96.28 - 96.71)		96.80 (96.59 - 97.00)	
2.00	80.21 (79.96 - 80.45)		40.45 (39.57 - 41.34)		93.11 (92.81 - 93.39)		93.80 (93.51 - 94.08)	
5.00	70.07 (69.70 - 70.45)		15.60 (14.87 - 16.35)		82.45 (81.94 - 82.94)		84.30 (83.91 - 84.68)	
10.00	56.00 (55.50 - 56.50)		4.77 (4.61 - 4.93)		64.53 (63.86 - 65.20)		68.69 (68.46 - 68.92)	
20.00	32.95 (32.39 - 33.52)		1.90 (1.84 - 1.96)		29.65 (28.81 - 30.50)		52.69 (51.85 - 53.53)	
50.00	1.38 (1.30 - 1.47)		0.27 (0.25 - 0.29)		5.96 (5.85 - 6.07)		29.82 (29.56 - 30.08)	

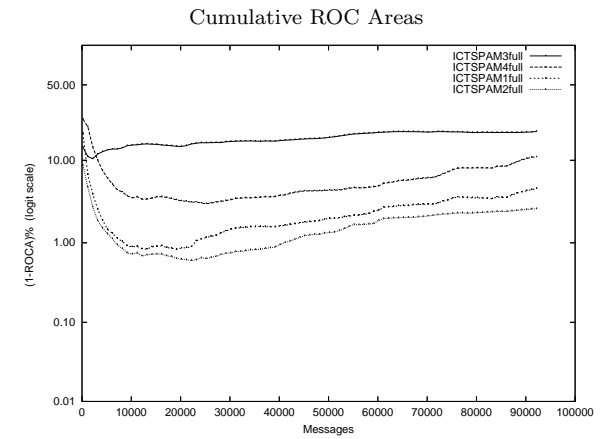
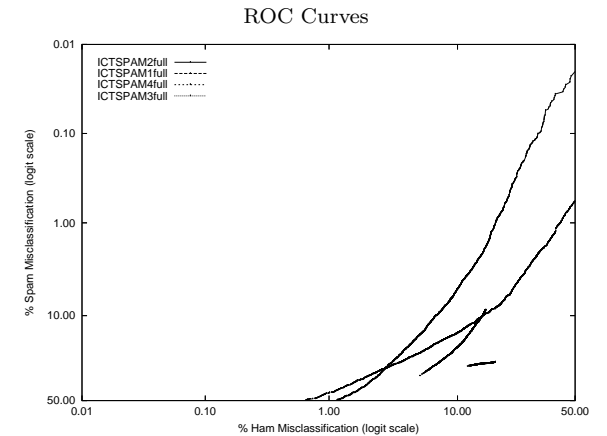


Learning Curves

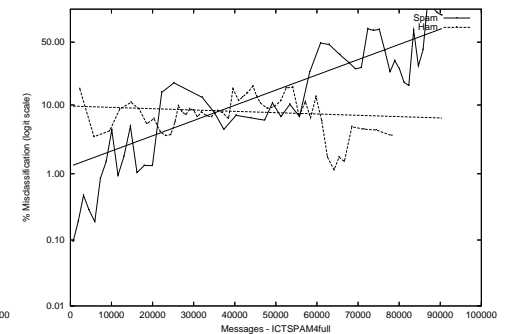
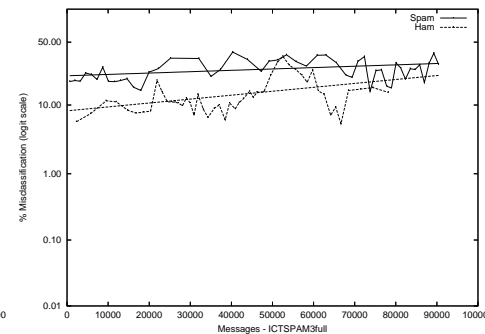
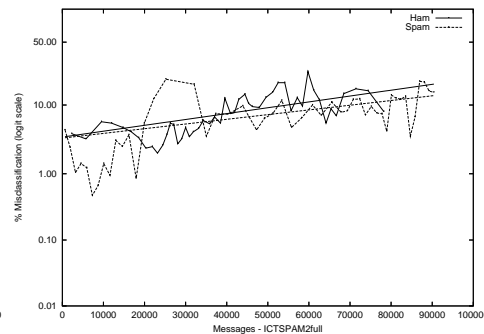
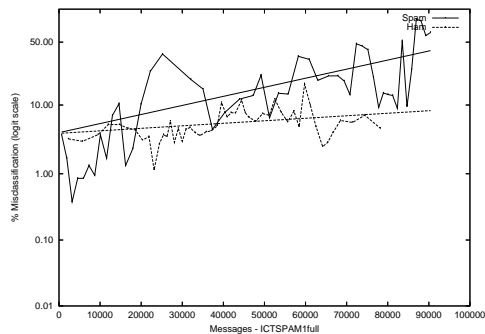


ICT - Chinese Academy of Sciences (ICT) - Public trec05p-1/full Corpus

Gold Std. Filter Classification	ICTSPAM1full			ICTSPAM2full			ICTSPAM3full			ICTSPAM4full		
	ham	spam	ham	spam	ham	spam	ham	spam	ham	spam		
Filter	ham	37158	11006	ham	36119	4241	ham	33842	14897	ham	36177	13141
Classification	spam	2241	41784	spam	3280	48549	spam	5557	37893	spam	3222	39649
Ham Misc%	5.69 (5.46-5.92)			8.33 (8.05-8.60)			14.10 (13.76-14.45)			8.18 (7.91-8.45)		
Spam Misc%	20.85 (20.50-21.20)			8.03 (7.80-8.27)			28.22 (27.84-28.61)			24.89 (24.52-25.26)		
Logit av Misc%	11.19 (10.96-11.43)			8.18 (8.00-8.36)			20.26 (19.99-20.54)			14.66 (14.41-14.92)		
(1-ROCA)%	4.659 (4.545 - 4.775)			2.643 (2.544 - 2.746)			20.485 (20.213 - 20.759)			10.952 (10.698 - 11.211)		
Ham Misc%	Spam Misc%											
0.01	96.34 (83.48 - 99.27)	91.73 (87.97 - 94.39)	99.95 (99.87 - 99.98)	99.89 (99.77 - 99.95)	99.89 (99.68 - 99.96)							
0.02	91.81 (84.62 - 95.81)	91.16 (89.74 - 92.41)	99.89 (99.77 - 99.95)	99.73 (99.47 - 99.86)	99.73 (99.47 - 99.86)							
0.05	80.47 (75.60 - 84.56)	88.29 (83.79 - 91.66)	99.71 (99.56 - 99.80)	99.31 (98.98 - 99.53)	99.31 (98.98 - 99.53)							
0.10	72.26 (66.40 - 77.44)	79.51 (76.50 - 82.22)	99.39 (99.14 - 99.57)	98.44 (97.95 - 98.82)	98.44 (97.95 - 98.82)							
0.20	61.57 (58.91 - 64.16)	72.22 (68.47 - 75.68)	98.68 (98.33 - 98.96)	97.30 (96.70 - 97.80)	97.30 (96.70 - 97.80)							
0.50	53.94 (52.25 - 55.62)	59.09 (57.96 - 60.20)	96.86 (96.34 - 97.31)	93.92 (93.08 - 94.67)	93.92 (93.08 - 94.67)							
1.00	45.19 (44.07 - 46.31)	52.25 (50.25 - 54.25)	94.10 (93.47 - 94.66)	87.77 (86.39 - 89.02)	87.77 (86.39 - 89.02)							
2.00	35.25 (34.05 - 36.48)	40.16 (38.17 - 42.18)	88.73 (87.86 - 89.55)	75.30 (73.53 - 76.99)	75.30 (73.53 - 76.99)							
5.00	23.21 (22.68 - 23.74)	16.23 (15.23 - 17.27)	70.40 (69.04 - 71.71)	37.11 (34.18 - 40.13)	37.11 (34.18 - 40.13)							
10.00	14.86 (14.42 - 15.31)	5.31 (4.99 - 5.64)	40.31 (38.49 - 42.16)	20.25 (19.52 - 21.00)	20.25 (19.52 - 21.00)							
20.00	7.64 (7.32 - 7.98)	0.77 (0.69 - 0.86)	26.68 (26.27 - 27.10)	11.95 (11.67 - 12.23)	11.95 (11.67 - 12.23)							
50.00	0.58 (0.52 - 0.65)	0.02 (0.01 - 0.04)	16.27 (15.91 - 16.63)	7.38 (7.14 - 7.63)	7.38 (7.14 - 7.63)							
Spam Misc%	Ham Misc%											
0.01	99.76 (97.89 - 99.97)	60.66 (51.46 - 69.16)	99.95 (99.86 - 99.98)	99.91 (99.78 - 99.96)								
0.02	97.53 (90.95 - 99.36)	50.92 (44.11 - 57.69)	99.90 (99.78 - 99.95)	99.89 (99.78 - 99.95)								
0.05	88.29 (80.89 - 93.07)	38.55 (34.24 - 43.04)	99.87 (99.79 - 99.91)	99.60 (99.39 - 99.73)								
0.10	79.61 (74.59 - 83.84)	34.07 (31.60 - 36.62)	99.71 (99.60 - 99.79)	99.42 (99.22 - 99.57)								
0.20	67.00 (64.76 - 69.17)	28.33 (26.69 - 30.03)	99.45 (99.31 - 99.57)	98.84 (98.59 - 99.04)								
0.50	52.26 (49.67 - 54.84)	22.58 (21.85 - 23.32)	98.56 (98.34 - 98.75)	97.00 (96.65 - 97.31)								
1.00	43.32 (42.11 - 44.53)	18.43 (17.90 - 18.98)	97.09 (96.77 - 97.38)	93.42 (92.81 - 93.97)								
2.00	34.74 (33.67 - 35.84)	15.37 (14.95 - 15.80)	93.91 (93.48 - 94.32)	86.90 (86.08 - 87.67)								
5.00	24.45 (23.92 - 24.99)	10.30 (10.01 - 10.61)	84.83 (84.14 - 85.50)	66.19 (64.81 - 67.54)								
10.00	15.23 (14.78 - 15.70)	6.94 (6.62 - 7.28)	69.55 (68.66 - 70.42)	33.24 (31.40 - 35.13)								
20.00	6.35 (6.09 - 6.61)	4.23 (4.01 - 4.47)	39.67 (38.55 - 40.80)	10.13 (9.79 - 10.47)								
50.00	0.65 (0.57 - 0.74)	1.16 (1.06 - 1.27)	8.28 (7.99 - 8.58)	4.01 (3.81 - 4.21)								

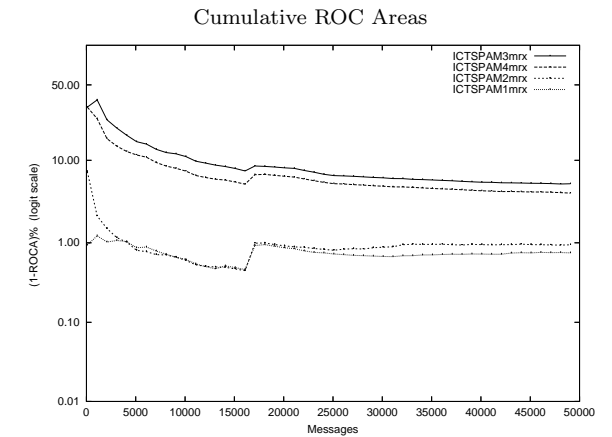
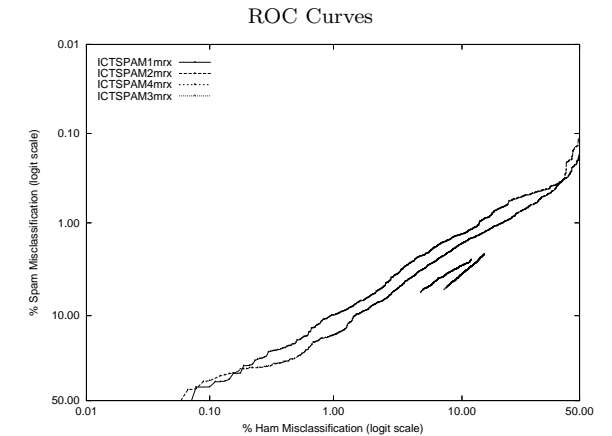


Learning Curves

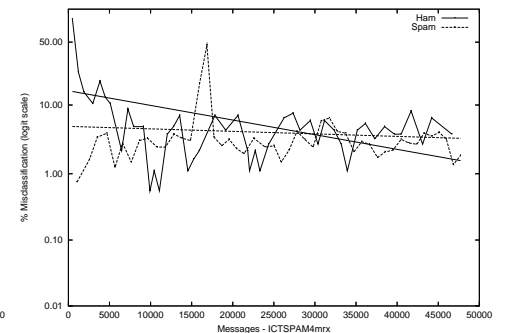
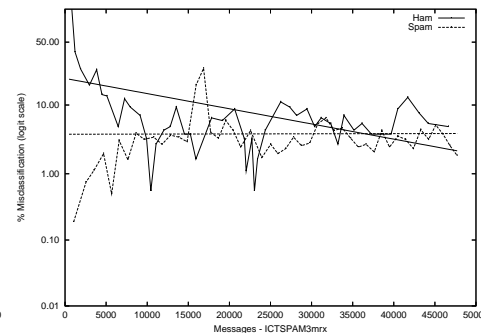
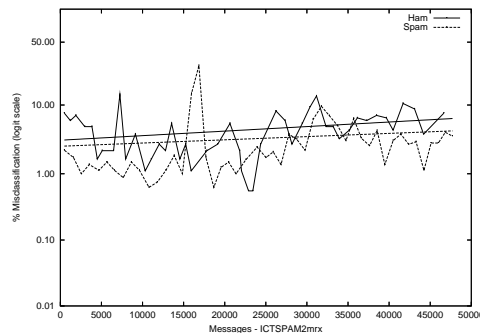
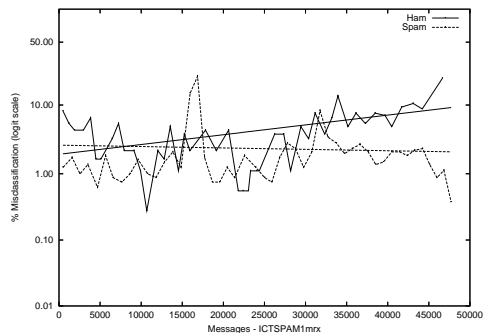


ICT - Chinese Academy of Sciences (ICT) - Private Mr X. Corpus

Gold Std. Filter Classification	ICTSPAM1mrx		ICTSPAM2mrx		ICTSPAM3mrx		ICTSPAM4mrx	
	ham	spam	ham	spam	ham	spam	ham	spam
Ham Misc%	8634	951	8630	1369	8173	1565	8434	1641
Spam Misc%	404	39097	408	38679	865	38483	604	38407
Logit av Misc%	4.47 (4.05-4.92)		4.51 (4.10-4.96)		9.57 (8.97-10.20)		6.68 (6.18-7.22)	
(1-ROCA)%	2.37 (2.23-2.53)		3.42 (3.24-3.60)		3.91 (3.72-4.10)		4.10 (3.91-4.30)	
Ham Misc%	3.26 (3.08-3.46)		3.93 (3.72-4.15)		6.16 (5.91-6.41)		5.24 (5.01-5.49)	
	0.748 (0.675 - 0.828)		0.943 (0.862 - 1.033)		5.328 (4.984 - 5.694)		4.114 (3.848 - 4.399)	
Ham Misc%	Spam Misc%							
0.01	98.95 (83.46 - 99.94)	88.32 (64.41 - 96.93)	99.92 (nan - nan)	99.92 (nan - nan)	99.92 (nan - nan)	99.92 (nan - nan)	99.92 (nan - nan)	99.92 (nan - nan)
0.02	91.21 (37.67 - 99.44)	69.93 (37.11 - 90.16)	99.55 (98.20 - 99.89)	99.55 (98.20 - 99.89)	99.55 (98.20 - 99.89)	99.55 (98.20 - 99.89)	99.55 (98.20 - 99.89)	99.55 (98.20 - 99.89)
0.05	68.91 (42.91 - 86.73)	53.50 (36.24 - 69.95)	99.34 (98.47 - 99.72)	99.34 (98.47 - 99.72)	99.34 (98.47 - 99.72)	99.34 (98.47 - 99.72)	99.34 (98.47 - 99.72)	99.34 (98.47 - 99.72)
0.10	41.24 (24.66 - 60.07)	37.43 (28.07 - 47.84)	98.50 (96.98 - 99.26)	98.50 (96.98 - 99.26)	98.50 (96.98 - 99.26)	98.50 (96.98 - 99.26)	98.50 (96.98 - 99.26)	98.50 (96.98 - 99.26)
0.20	28.92 (21.58 - 37.56)	30.54 (27.43 - 33.84)	97.33 (95.66 - 98.36)	97.33 (95.66 - 98.36)	97.33 (95.66 - 98.36)	97.33 (95.66 - 98.36)	97.33 (95.66 - 98.36)	97.33 (95.66 - 98.36)
0.50	18.19 (14.55 - 22.51)	25.24 (20.64 - 30.48)	93.85 (91.60 - 95.54)	93.85 (91.60 - 95.54)	93.85 (91.60 - 95.54)	93.85 (91.60 - 95.54)	93.85 (91.60 - 95.54)	93.85 (91.60 - 95.54)
1.00	9.89 (8.67 - 11.26)	15.64 (14.09 - 17.32)	87.99 (84.76 - 90.61)	87.99 (84.76 - 90.61)	87.99 (84.76 - 90.61)	87.99 (84.76 - 90.61)	87.99 (84.76 - 90.61)	87.99 (84.76 - 90.61)
2.00	6.18 (5.41 - 7.04)	8.11 (7.09 - 9.26)	74.89 (70.85 - 78.53)	74.89 (70.85 - 78.53)	74.89 (70.85 - 78.53)	74.89 (70.85 - 78.53)	74.89 (70.85 - 78.53)	74.89 (70.85 - 78.53)
5.00	2.32 (2.12 - 2.55)	3.32 (2.97 - 3.71)	36.85 (30.79 - 43.36)	36.85 (30.79 - 43.36)	36.85 (30.79 - 43.36)	36.85 (30.79 - 43.36)	36.85 (30.79 - 43.36)	36.85 (30.79 - 43.36)
10.00	1.34 (1.24 - 1.45)	1.71 (1.55 - 1.87)	3.70 (3.37 - 4.06)	3.70 (3.37 - 4.06)	3.70 (3.37 - 4.06)	3.70 (3.37 - 4.06)	3.70 (3.37 - 4.06)	3.70 (3.37 - 4.06)
20.00	0.65 (0.58 - 0.74)	0.96 (0.87 - 1.07)	2.46 (2.32 - 2.61)	2.46 (2.32 - 2.61)	2.46 (2.32 - 2.61)	2.46 (2.32 - 2.61)	2.46 (2.32 - 2.61)	2.46 (2.32 - 2.61)
50.00	0.18 (0.14 - 0.23)	0.11 (0.08 - 0.16)	1.56 (1.46 - 1.67)	1.56 (1.46 - 1.67)	1.56 (1.46 - 1.67)	1.56 (1.46 - 1.67)	1.56 (1.46 - 1.67)	1.56 (1.46 - 1.67)
Spam Misc%	Ham Misc%							
0.01	87.46 (79.63 - 92.57)	82.97 (80.59 - 85.11)	99.58 (98.64 - 99.87)	99.58 (98.64 - 99.87)	99.58 (98.64 - 99.87)	99.58 (98.64 - 99.87)	99.58 (98.64 - 99.87)	99.58 (98.64 - 99.87)
0.02	86.57 (77.92 - 92.17)	81.93 (74.58 - 87.52)	99.31 (98.41 - 99.71)	99.31 (98.41 - 99.71)	99.31 (98.41 - 99.71)	99.31 (98.41 - 99.71)	99.31 (98.41 - 99.71)	99.31 (98.41 - 99.71)
0.05	64.30 (54.40 - 73.10)	61.48 (51.29 - 70.77)	98.51 (97.64 - 99.06)	98.51 (97.64 - 99.06)	98.51 (97.64 - 99.06)	98.51 (97.64 - 99.06)	98.51 (97.64 - 99.06)	98.51 (97.64 - 99.06)
0.10	60.17 (57.03 - 63.23)	51.16 (47.28 - 55.03)	97.08 (95.94 - 97.91)	97.08 (95.94 - 97.91)	97.08 (95.94 - 97.91)	97.08 (95.94 - 97.91)	97.08 (95.94 - 97.91)	97.08 (95.94 - 97.91)
0.20	49.36 (46.89 - 51.83)	45.51 (43.07 - 47.97)	93.45 (91.69 - 94.86)	93.45 (91.69 - 94.86)	93.45 (91.69 - 94.86)	93.45 (91.69 - 94.86)	93.45 (91.69 - 94.86)	93.45 (91.69 - 94.86)
0.50	26.28 (20.51 - 32.99)	35.05 (31.50 - 38.77)	85.28 (83.11 - 87.22)	85.28 (83.11 - 87.22)	85.28 (83.11 - 87.22)	85.28 (83.11 - 87.22)	85.28 (83.11 - 87.22)	85.28 (83.11 - 87.22)
1.00	13.35 (12.39 - 14.39)	19.38 (17.19 - 21.78)	68.99 (65.81 - 71.99)	68.99 (65.81 - 71.99)	68.99 (65.81 - 71.99)	68.99 (65.81 - 71.99)	68.99 (65.81 - 71.99)	68.99 (65.81 - 71.99)
2.00	5.91 (5.28 - 6.60)	8.55 (7.72 - 9.47)	35.56 (31.58 - 39.75)	35.56 (31.58 - 39.75)	35.56 (31.58 - 39.75)	35.56 (31.58 - 39.75)	35.56 (31.58 - 39.75)	35.56 (31.58 - 39.75)
5.00	2.52 (2.20 - 2.89)	3.37 (2.96 - 3.85)	7.84 (7.26 - 8.47)	7.84 (7.26 - 8.47)	7.84 (7.26 - 8.47)	7.84 (7.26 - 8.47)	7.84 (7.26 - 8.47)	7.84 (7.26 - 8.47)
10.00	0.96 (0.74 - 1.25)	1.52 (1.26 - 1.82)	7.03 (6.49 - 7.61)	7.03 (6.49 - 7.61)	7.03 (6.49 - 7.61)	7.03 (6.49 - 7.61)	7.03 (6.49 - 7.61)	7.03 (6.49 - 7.61)
20.00	0.42 (0.30 - 0.58)	0.65 (0.49 - 0.87)	6.23 (5.71 - 6.79)	6.23 (5.71 - 6.79)	6.23 (5.71 - 6.79)	6.23 (5.71 - 6.79)	6.23 (5.71 - 6.79)	6.23 (5.71 - 6.79)
50.00	0.08 (0.04 - 0.17)	0.07 (0.03 - 0.16)	4.02 (3.59 - 4.49)	4.02 (3.59 - 4.49)	4.02 (3.59 - 4.49)	4.02 (3.59 - 4.49)	4.02 (3.59 - 4.49)	4.02 (3.59 - 4.49)

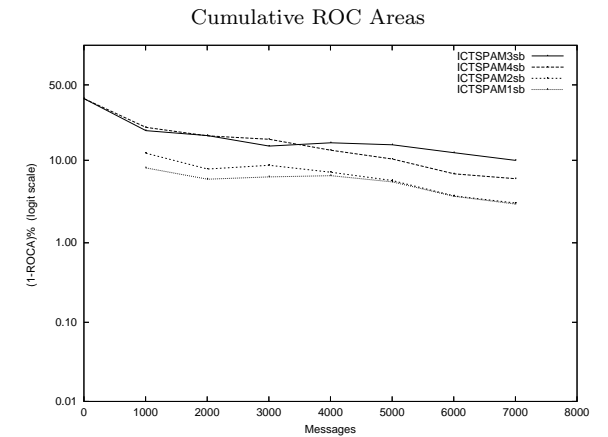
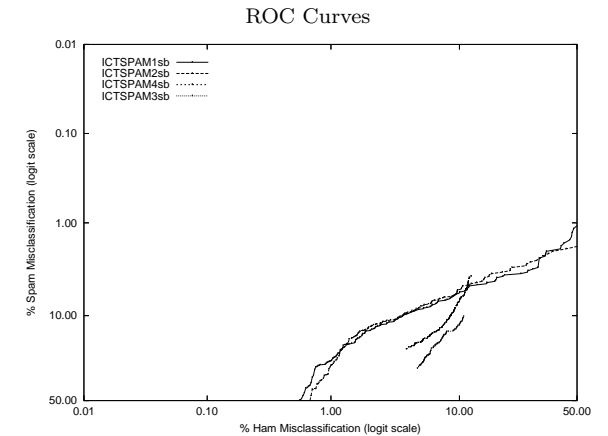


Learning Curves

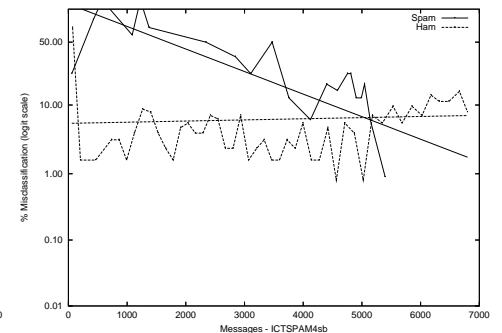
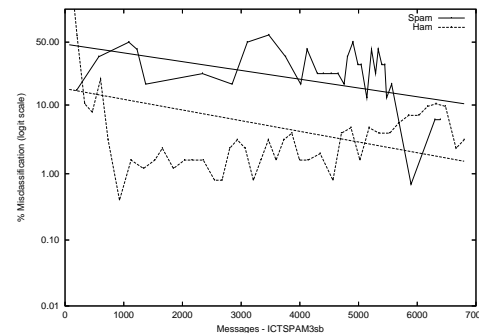
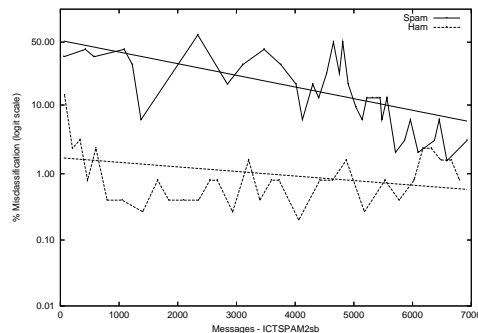
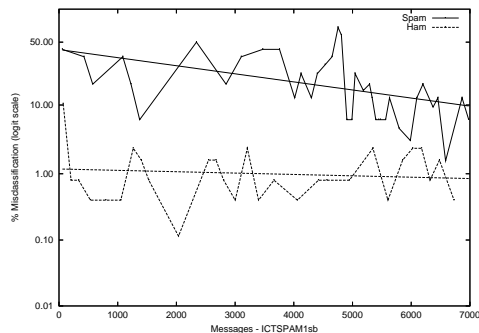


ICT - Chinese Academy of Sciences (ICT) - Private S.B. Corpus

Gold Std. Filter Classification	ICTSPAM1sb		ICTSPAM2sb		ICTSPAM3sb		ICTSPAM4sb	
	ham	spam	ham	spam	ham	spam	ham	spam
ham	6168	140	6164	122	5819	150	5838	116
spam	63	635	67	653	412	625	393	659
Ham Misc%	1.01 (0.78-1.29)		1.08 (0.83-1.36)		6.61 (6.01-7.26)		6.31 (5.72-6.94)	
Spam Misc%	18.06 (15.42-20.96)		15.74 (13.25-18.50)		19.35 (16.63-22.32)		14.97 (12.53-17.68)	
Logit av Misc%	4.53 (3.91-5.25)		4.31 (3.72-5.00)		11.53 (10.53-12.62)		9.82 (8.88-10.85)	
(1-ROCA)%	3.023 (2.338 - 3.900)		3.110 (2.429 - 3.975)		9.985 (8.711 - 11.423)		6.112 (5.139 - 7.256)	
Ham Misc%	Spam Misc%							
0.01	99.87 (nan - nan)		99.87 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	99.74 (nan - nan)		99.87 (nan - nan)		99.48 (nan - nan)		99.35 (nan - nan)	
0.05	99.74 (nan - nan)		99.74 (nan - nan)		98.84 (nan - nan)		98.32 (94.23 - 99.53)	
0.10	97.55 (nan - nan)		99.35 (nan - nan)		98.71 (nan - nan)		97.03 (92.86 - 98.80)	
0.20	92.39 (66.12 - 98.69)		97.29 (75.38 - 99.76)		97.42 (93.92 - 98.93)		95.74 (92.38 - 97.66)	
0.50	51.87 (36.43 - 66.96)		63.48 (43.62 - 79.62)		94.71 (91.95 - 96.56)		90.19 (85.47 - 93.50)	
1.00	26.45 (20.73 - 33.09)		28.77 (19.00 - 41.03)		89.03 (85.48 - 91.80)		79.10 (72.88 - 84.20)	
2.00	14.19 (11.47 - 17.43)		13.29 (10.82 - 16.22)		74.06 (69.16 - 78.43)		62.84 (54.67 - 70.34)	
5.00	8.65 (6.74 - 11.02)		8.39 (6.50 - 10.75)		29.03 (24.15 - 34.45)		18.32 (15.77 - 21.19)	
10.00	5.94 (4.32 - 8.10)		5.42 (3.89 - 7.50)		13.16 (10.98 - 15.69)		7.10 (5.31 - 9.43)	
20.00	4.00 (2.76 - 5.76)		3.48 (2.38 - 5.07)		11.10 (9.14 - 13.41)		5.94 (4.47 - 7.84)	
50.00	1.16 (0.55 - 2.44)		1.94 (1.18 - 3.15)		6.45 (5.07 - 8.18)		3.61 (2.52 - 5.16)	
Spam Misc%	Ham Misc%							
0.01	-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)	
0.02	-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)	
0.05	-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)		-100.00 (nan - nan)	
0.10	93.28 (85.32 - 97.07)		98.59 (94.50 - 99.65)		99.25 (nan - nan)		97.21 (76.97 - 99.73)	
0.20	85.41 (70.79 - 93.40)		96.79 (83.24 - 99.46)		98.49 (88.28 - 99.82)		95.59 (82.57 - 99.00)	
0.50	85.19 (68.49 - 93.83)		90.32 (61.67 - 98.19)		97.98 (93.40 - 99.40)		90.37 (75.96 - 96.54)	
1.00	63.76 (37.77 - 83.61)		69.92 (49.45 - 84.68)		95.51 (90.40 - 97.96)		84.50 (71.95 - 92.05)	
2.00	42.34 (30.37 - 55.27)		44.94 (22.59 - 69.53)		87.66 (79.80 - 92.74)		75.24 (59.97 - 86.04)	
5.00	11.75 (6.60 - 20.05)		10.54 (6.64 - 16.35)		68.27 (57.18 - 77.61)		28.21 (12.63 - 51.66)	
10.00	3.76 (2.56 - 5.49)		3.74 (2.45 - 5.68)		25.53 (14.52 - 40.90)		8.70 (7.61 - 9.93)	
20.00	1.28 (0.96 - 1.71)		1.25 (0.98 - 1.59)		6.42 (5.53 - 7.44)		4.25 (3.42 - 5.27)	
50.00	0.56 (0.39 - 0.81)		0.69 (0.49 - 0.97)		3.53 (3.12 - 3.99)		2.68 (2.27 - 3.16)	

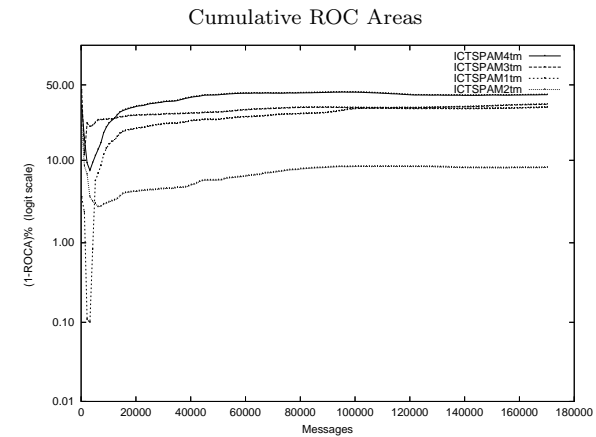
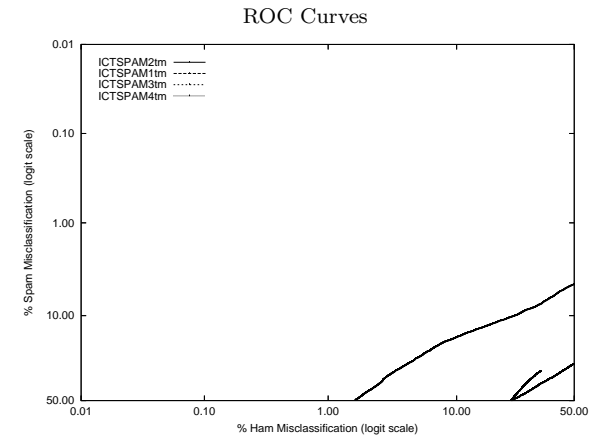


Learning Curves

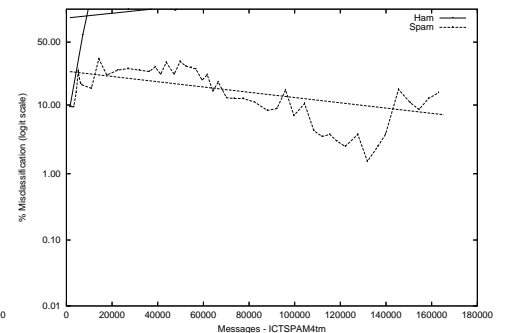
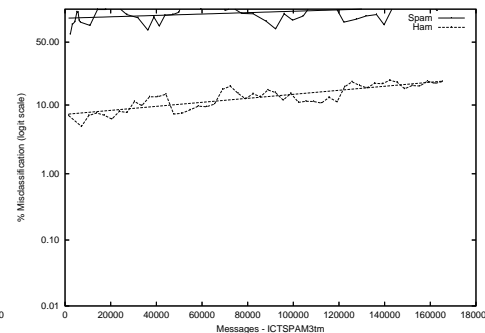
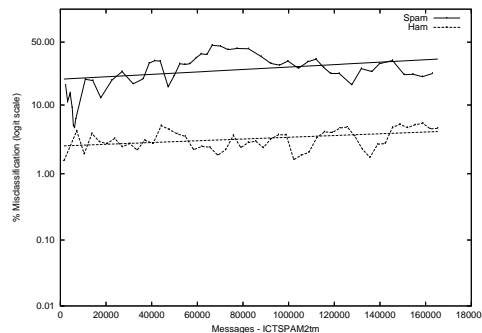
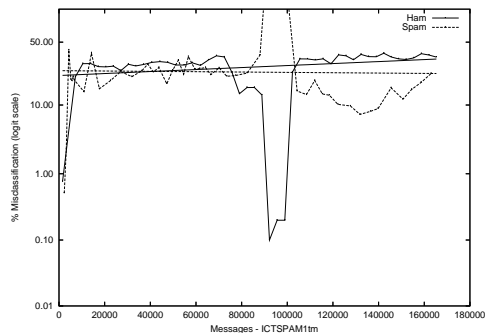


ICT - Chinese Academy of Sciences (ICT) - Private T.M. Corpus

Gold Std.	ICTSPAM1tm			ICTSPAM2tm			ICTSPAM3tm			ICTSPAM4tm			
	Filter	ham	spam	ham	spam	ham	spam	ham	spam	ham	spam		
Classification	ham spam	105836 44849	5098 14418	ham spam	145586 5099	5350 14166	ham spam	130602 20083	14304 5212	ham spam	27300 123385	3228 16288	
Ham Misc%		29.76 (29.53-29.99)			3.38 (3.29-3.48)			13.33 (13.16-13.50)			81.88 (81.69-82.08)		
Spam Misc%		26.12 (25.51-26.74)			27.41 (26.79-28.05)			73.29 (72.67-73.91)			16.54 (16.02-17.07)		
Logit av Misc%		27.91 (27.57-28.25)			10.31 (10.12-10.51)			39.38 (38.96-39.80)			48.62 (48.12-49.12)		
(1-ROCA)%		34.208 (33.769 - 34.649)			8.298 (8.055 - 8.549)			36.233 (35.844 - 36.623)			42.893 (42.412 - 43.376)		
Ham Misc%		Spam Misc%											
0.01		99.68 (99.37 - 99.84)			95.00 (92.87 - 96.51)			99.98 (nan - nan)			99.98 (nan - nan)		
0.02		99.39 (99.11 - 99.58)			94.03 (93.01 - 94.90)			99.93 (99.83 - 99.97)			99.97 (99.92 - 99.99)		
0.05		98.76 (98.46 - 99.00)			89.81 (88.26 - 91.17)			99.85 (99.75 - 99.91)			99.94 (99.87 - 99.97)		
0.10		98.13 (97.87 - 98.37)			86.36 (84.59 - 87.95)			99.75 (99.65 - 99.82)			99.83 (99.73 - 99.89)		
0.20		97.09 (96.71 - 97.44)			79.02 (77.64 - 80.34)			99.61 (99.51 - 99.69)			99.69 (99.59 - 99.76)		
0.50		94.97 (94.62 - 95.30)			71.36 (70.40 - 72.30)			99.04 (98.89 - 99.17)			99.23 (99.08 - 99.35)		
1.00		92.21 (91.77 - 92.62)			59.88 (58.57 - 61.18)			97.86 (97.63 - 98.06)			98.65 (98.46 - 98.81)		
2.00		88.12 (87.63 - 88.60)			45.05 (43.85 - 46.26)			95.65 (95.34 - 95.95)			97.38 (97.15 - 97.59)		
5.00		79.10 (78.51 - 79.68)			25.18 (24.43 - 25.95)			88.73 (88.25 - 89.19)			93.43 (93.03 - 93.80)		
10.00		69.05 (68.42 - 69.66)			16.33 (15.81 - 16.85)			78.78 (78.16 - 79.38)			87.06 (86.51 - 87.58)		
20.00		54.47 (53.77 - 55.18)			11.32 (10.86 - 11.80)			58.42 (57.59 - 59.24)			74.62 (73.92 - 75.30)		
50.00		28.00 (27.35 - 28.65)			4.72 (4.49 - 4.96)			31.51 (30.90 - 32.13)			35.45 (34.63 - 36.28)		
Spam Misc%		Ham Misc%											
0.01		100.00 (nan - nan)			99.56 (99.04 - 99.80)			99.99 (99.97 - 100.00)			99.99 (nan - nan)		
0.02		100.00 (nan - nan)			99.46 (98.89 - 99.74)			99.99 (99.96 - 100.00)			99.99 (99.97 - 100.00)		
0.05		100.00 (nan - nan)			98.77 (98.07 - 99.22)			99.92 (99.84 - 99.96)			99.97 (99.95 - 99.99)		
0.10		100.00 (nan - nan)			98.21 (97.36 - 98.79)			99.82 (99.71 - 99.89)			99.93 (99.89 - 99.95)		
0.20		100.00 (nan - nan)			96.99 (96.58 - 97.35)			99.71 (99.60 - 99.79)			99.89 (99.85 - 99.92)		
0.50		99.94 (99.83 - 99.98)			95.91 (95.67 - 96.15)			99.35 (99.21 - 99.46)			99.66 (99.58 - 99.72)		
1.00		99.40 (99.27 - 99.51)			94.55 (93.81 - 95.20)			98.61 (98.42 - 98.78)			99.29 (99.18 - 99.39)		
2.00		97.90 (97.35 - 98.33)			89.06 (87.57 - 90.40)			96.84 (96.51 - 97.15)			98.52 (98.34 - 98.68)		
5.00		94.79 (94.34 - 95.20)			47.68 (45.72 - 49.64)			92.37 (91.92 - 92.80)			96.47 (96.20 - 96.72)		
10.00		93.26 (93.11 - 93.41)			24.99 (23.35 - 26.70)			84.30 (83.61 - 84.98)			93.03 (92.69 - 93.36)		
20.00		63.73 (62.32 - 65.12)			6.94 (6.60 - 7.30)			68.57 (67.68 - 69.44)			62.88 (60.50 - 65.19)		
50.00		23.94 (23.24 - 24.65)			1.65 (1.57 - 1.73)			23.50 (23.12 - 23.88)			38.79 (38.12 - 39.46)		



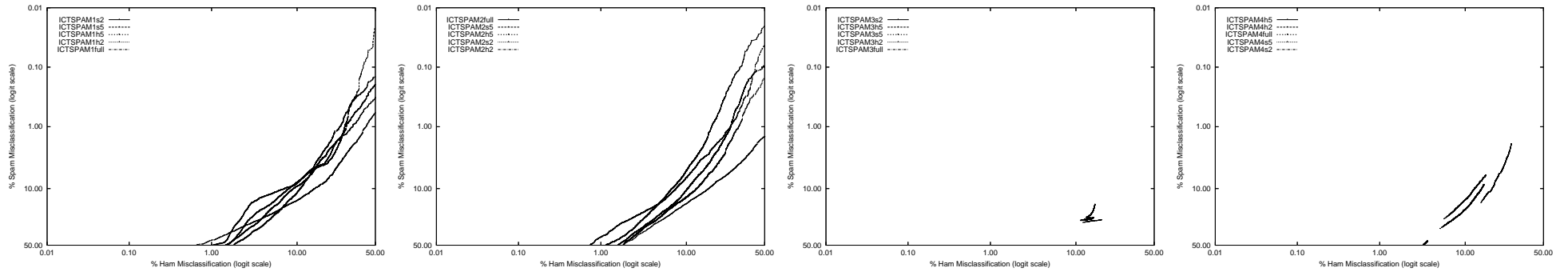
Learning Curves



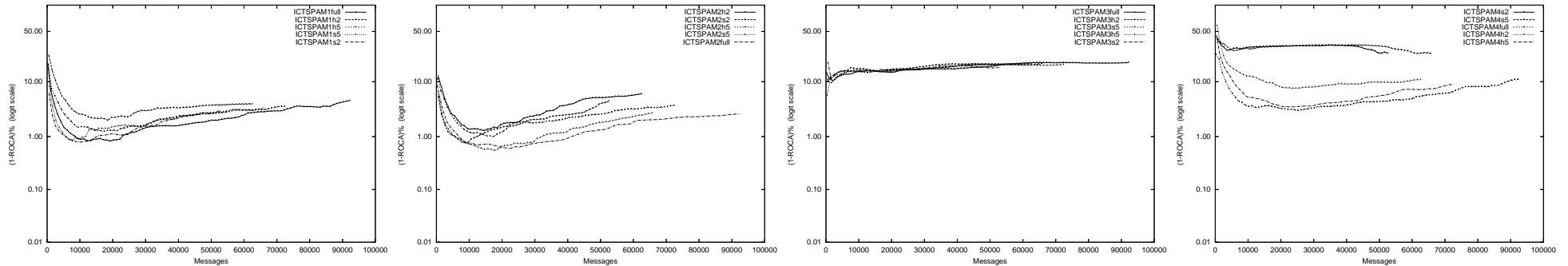
ICT - Chinese Academy of Sciences (ICT) - Public subset Corpora

	ICTSPAM1full	ICTSPAM1h2	ICTSPAM1h5	ICTSPAM1s2	ICTSPAM1s5
Ham Misc%	5.69 (5.46-5.92)	15.15 (14.44-15.87)	10.64 (10.21-11.08)	3.01 (2.84-3.18)	6.05 (5.82-6.29)
Spam Misc%	20.85 (20.50-21.20)	3.74 (3.58-3.90)	8.51 (8.27-8.75)	16.37 (15.74-17.01)	13.75 (13.34-14.18)
Logit av Misc%	11.19 (10.96-11.43)	7.69 (7.44-7.94)	9.52 (9.29-9.76)	7.23 (6.99-7.48)	9.20 (8.98-9.43)
(1-ROCA)%	4.659 (4.545 - 4.775)	4.012 (3.800 - 4.235)	3.611 (3.472 - 3.757)	3.036 (2.883 - 3.197)	3.325 (3.195 - 3.461)
	ICTSPAM2full	ICTSPAM2h2	ICTSPAM2h5	ICTSPAM2s2	ICTSPAM2s5
Ham Misc%	8.33 (8.05-8.60)	11.32 (10.70-11.97)	7.73 (7.36-8.11)	6.91 (6.66-7.16)	5.57 (5.35-5.80)
Spam Misc%	8.03 (7.80-8.27)	14.29 (13.99-14.59)	15.66 (15.35-15.97)	17.98 (17.33-18.65)	15.47 (15.04-15.92)
Logit av Misc%	8.18 (8.00-8.36)	12.73 (12.36-13.11)	11.09 (10.81-11.38)	11.31 (11.02-11.61)	9.42 (9.19-9.65)
(1-ROCA)%	2.643 (2.544 - 2.746)	6.140 (5.862 - 6.431)	3.777 (3.622 - 3.938)	4.571 (4.409 - 4.738)	2.741 (2.648 - 2.837)
	ICTSPAM3full	ICTSPAM3h2	ICTSPAM3h5	ICTSPAM3s2	ICTSPAM3s5
Ham Misc%	14.10 (13.76-14.45)	13.68 (13.00-14.38)	12.51 (12.05-12.99)	14.42 (14.08-14.78)	13.50 (13.16-13.84)
Spam Misc%	28.22 (27.84-28.61)	26.99 (26.61-27.37)	27.21 (26.83-27.59)	23.67 (22.94-24.40)	27.17 (26.64-27.72)
Logit av Misc%	20.26 (19.99-20.54)	19.49 (19.01-19.97)	18.78 (18.43-19.14)	18.61 (18.24-18.98)	19.44 (19.13-19.75)
(1-ROCA)%	20.485 (20.213 - 20.759)	19.947 (19.538 - 20.363)	19.044 (18.683 - 19.411)	17.086 (16.673 - 17.507)	19.558 (19.192 - 19.930)
	ICTSPAM4full	ICTSPAM4h2	ICTSPAM4h5	ICTSPAM4s2	ICTSPAM4s5
Ham Misc%	8.18 (7.91-8.45)	19.51 (18.72-20.31)	8.31 (7.93-8.71)	1.60 (1.48-1.73)	1.60 (1.48-1.73)
Spam Misc%	24.89 (24.52-25.26)	9.65 (9.40-9.91)	18.46 (18.13-18.80)	64.28 (63.46-65.10)	64.24 (63.66-64.82)
Logit av Misc%	14.66 (14.41-14.92)	13.86 (13.52-14.21)	12.53 (12.23-12.84)	14.61 (14.08-15.16)	14.60 (14.10-15.13)
(1-ROCA)%	10.952 (10.698 - 11.211)	10.821 (10.387 - 11.271)	8.995 (8.751 - 9.247)	27.891 (27.291 - 28.498)	27.506 (27.078 - 27.939)

ROC Curves

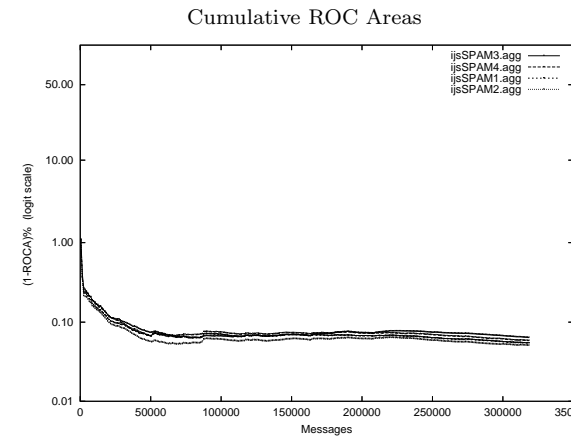
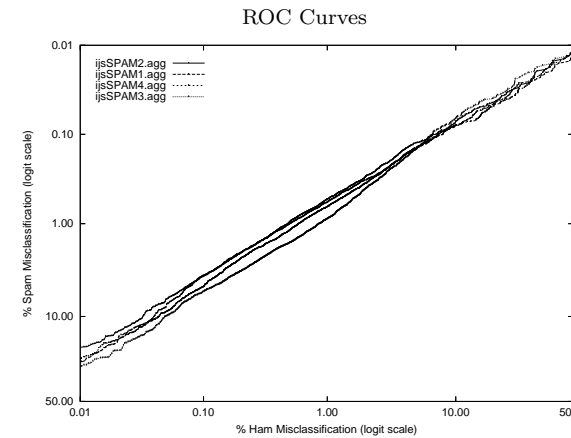


Cumulative ROC Area

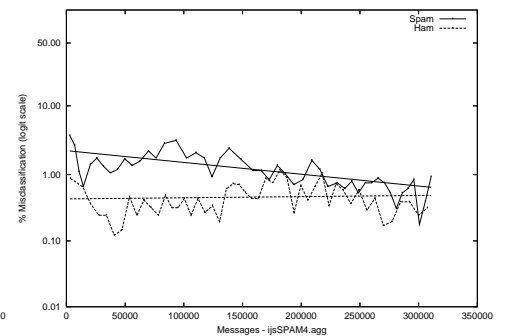
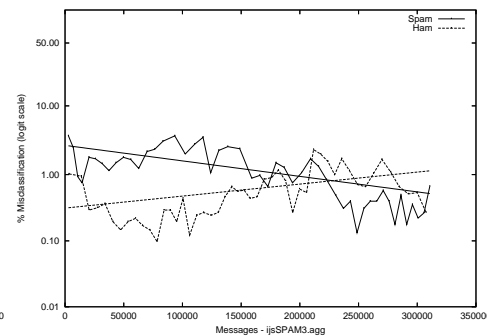
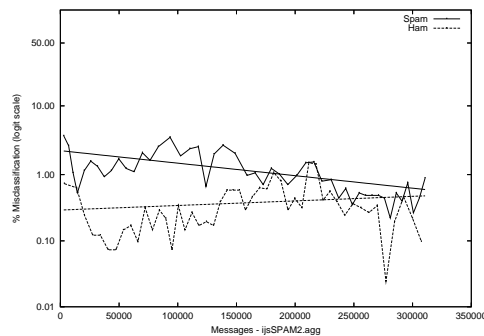
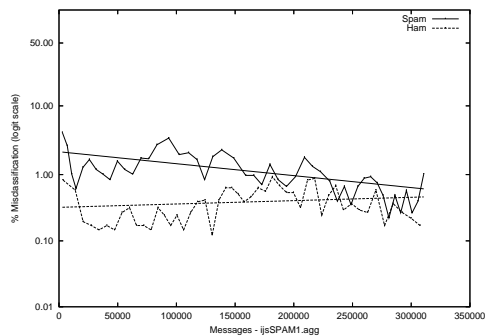


ijs - Jozef Stefan Inst. on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	ijsSPAM1.agg		ijsSPAM2.agg		ijsSPAM3.agg		ijsSPAM4.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	204557	1381	204575	1400	204031	1488	204411	1444
	796	111748	778	111729	1322	111641	941	111685
Ham Misc%	0.39 (0.36-0.42)		0.38 (0.35-0.41)		0.64 (0.61-0.68)		0.46 (0.43-0.49)	
Spam Misc%	1.22 (1.16-1.29)		1.24 (1.17-1.30)		1.32 (1.25-1.38)		1.28 (1.21-1.34)	
Logit av Misc%	0.69 (0.66-0.72)		0.69 (0.66-0.72)		0.92 (0.89-0.96)		0.77 (0.73-0.80)	
(1-ROCA)%	0.054 (0.048 - 0.061)		0.051 (0.044 - 0.058)		0.064 (0.058 - 0.070)		0.058 (0.052 - 0.065)	
Ham Misc%	Spam Misc%							
0.01	26.92 (21.78 - 32.76)		19.81 (17.01 - 22.95)		28.99 (23.99 - 34.56)		24.99 (18.83 - 32.35)	
0.02	17.78 (13.52 - 23.02)		14.32 (11.78 - 17.29)		21.17 (16.78 - 26.34)		17.02 (14.58 - 19.77)	
0.05	8.05 (6.52 - 9.89)		6.65 (5.57 - 7.91)		9.95 (8.09 - 12.19)		9.01 (7.53 - 10.75)	
0.10	3.73 (3.35 - 4.14)		3.78 (3.36 - 4.25)		5.54 (4.99 - 6.16)		4.91 (4.18 - 5.76)	
0.20	2.12 (1.93 - 2.32)		2.11 (1.94 - 2.30)		3.35 (3.12 - 3.59)		2.44 (2.19 - 2.71)	
0.50	1.01 (0.94 - 1.08)		0.95 (0.85 - 1.06)		1.62 (1.53 - 1.72)		1.19 (1.09 - 1.29)	
1.00	0.58 (0.53 - 0.63)		0.53 (0.49 - 0.59)		0.88 (0.82 - 0.95)		0.65 (0.60 - 0.71)	
2.00	0.33 (0.30 - 0.36)		0.31 (0.27 - 0.35)		0.42 (0.38 - 0.46)		0.36 (0.32 - 0.40)	
5.00	0.15 (0.13 - 0.17)		0.13 (0.11 - 0.15)		0.15 (0.13 - 0.18)		0.14 (0.13 - 0.17)	
10.00	0.08 (0.06 - 0.10)		0.07 (0.06 - 0.09)		0.07 (0.05 - 0.08)		0.08 (0.07 - 0.10)	
20.00	0.04 (0.03 - 0.06)		0.04 (0.03 - 0.06)		0.04 (0.03 - 0.05)		0.05 (0.04 - 0.06)	
50.00	0.01 (0.01 - 0.02)		0.01 (0.01 - 0.02)		0.01 (0.01 - 0.02)		0.01 (0.01 - 0.02)	
Spam Misc%	Ham Misc%							
0.01	52.33 (41.03 - 63.39)		52.20 (32.11 - 71.61)		64.01 (42.86 - 80.84)		54.64 (39.05 - 69.36)	
0.02	40.09 (28.83 - 52.51)		33.67 (23.46 - 45.67)		32.52 (19.09 - 49.60)		36.63 (27.31 - 47.08)	
0.05	17.00 (12.62 - 22.50)		14.84 (9.93 - 21.58)		13.41 (10.06 - 17.65)		17.56 (13.27 - 22.88)	
0.10	8.07 (6.34 - 10.21)		7.15 (5.55 - 9.17)		6.75 (5.91 - 7.70)		7.41 (6.28 - 8.74)	
0.20	3.58 (3.15 - 4.06)		3.15 (2.72 - 3.64)		3.85 (3.45 - 4.30)		3.62 (3.19 - 4.11)	
0.50	1.19 (1.07 - 1.32)		1.09 (0.96 - 1.23)		1.71 (1.58 - 1.86)		1.37 (1.23 - 1.52)	
1.00	0.50 (0.47 - 0.54)		0.48 (0.44 - 0.53)		0.86 (0.79 - 0.94)		0.59 (0.55 - 0.64)	
2.00	0.21 (0.19 - 0.23)		0.21 (0.19 - 0.24)		0.37 (0.34 - 0.40)		0.25 (0.23 - 0.28)	
5.00	0.07 (0.06 - 0.09)		0.07 (0.06 - 0.08)		0.12 (0.10 - 0.13)		0.10 (0.08 - 0.11)	
10.00	0.04 (0.03 - 0.05)		0.03 (0.03 - 0.04)		0.05 (0.04 - 0.06)		0.04 (0.04 - 0.06)	
20.00	0.02 (0.01 - 0.02)		0.01 (0.01 - 0.02)		0.02 (0.02 - 0.03)		0.01 (0.01 - 0.02)	
50.00	0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.00 (0.00 - 0.01)		0.00 (0.00 - 0.01)	

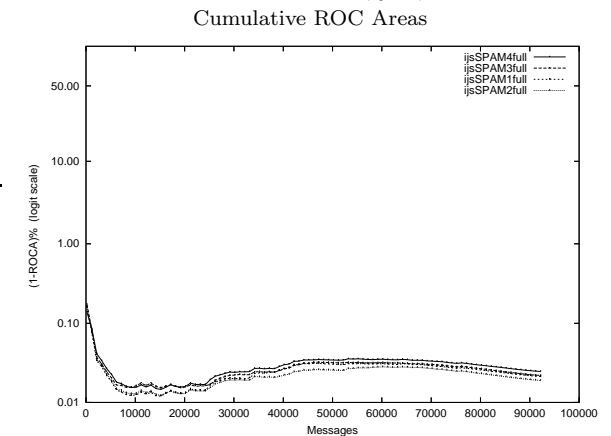
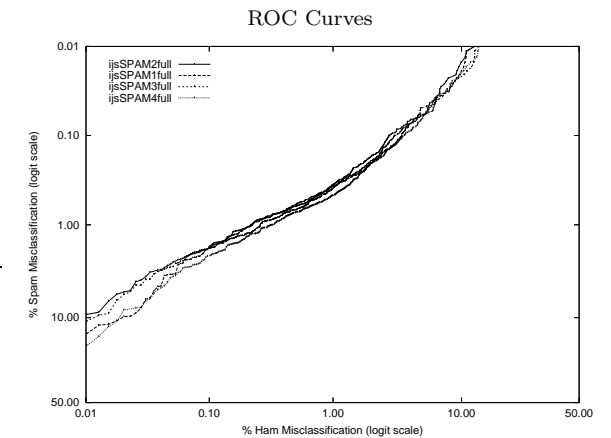


Learning Curves

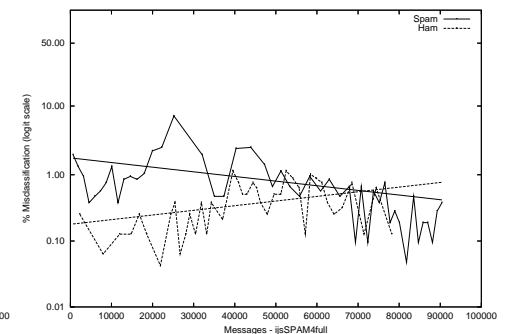
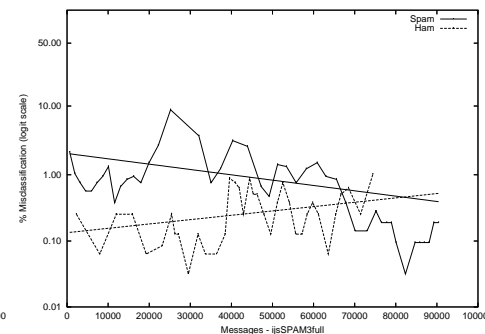
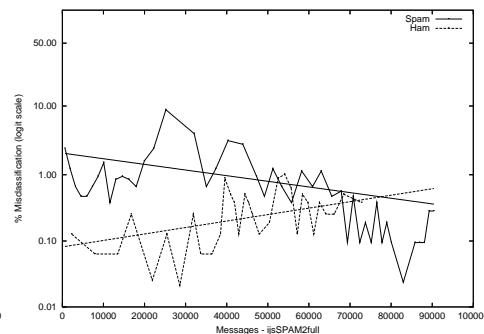
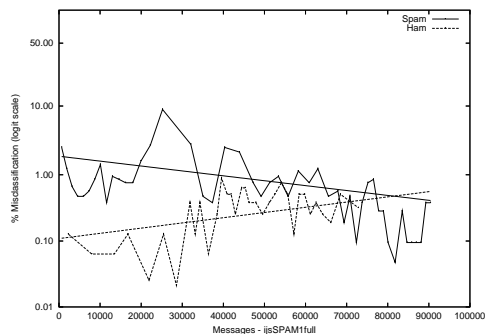


ijs - Jozef Stefan Inst. on Public trec05p-1/full Corpus

Gold Std. Filter Classification	ijsSPAM1full		ijsSPAM2full		ijsSPAM3full		ijsSPAM4full	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	39302 97	492 52298	39309 90	503 52287	39296 103	514 52276	39254 145	479 52311
Ham Misc%	0.25 (0.20-0.30)		0.23 (0.18-0.28)		0.26 (0.21-0.32)		0.37 (0.31-0.43)	
Spam Misc%	0.93 (0.85-1.02)		0.95 (0.87-1.04)		0.97 (0.89-1.06)		0.91 (0.83-0.99)	
Logit av Misc%	0.48 (0.43-0.53)		0.47 (0.42-0.52)		0.51 (0.45-0.56)		0.58 (0.53-0.63)	
(1-ROCA)%	0.021 (0.018 - 0.025)		0.019 (0.015 - 0.023)		0.022 (0.019 - 0.026)		0.025 (0.021 - 0.029)	
Ham Misc%	Spam Misc%							
0.01	16.22 (10.73 - 23.77)		9.89 (5.01 - 18.60)		11.97 (5.16 - 25.36)		19.10 (11.89 - 29.22)	
0.02	10.75 (7.42 - 15.33)		5.72 (2.89 - 11.00)		6.64 (3.22 - 13.19)		10.80 (5.51 - 20.10)	
0.05	3.52 (2.09 - 5.89)		2.73 (1.92 - 3.86)		2.95 (2.09 - 4.14)		3.84 (2.24 - 6.49)	
0.10	1.84 (1.41 - 2.38)		1.78 (1.42 - 2.22)		1.84 (1.55 - 2.18)		2.22 (1.74 - 2.83)	
0.20	1.08 (0.88 - 1.34)		1.09 (0.92 - 1.29)		1.27 (1.10 - 1.47)		1.36 (1.15 - 1.61)	
0.50	0.61 (0.53 - 0.70)		0.63 (0.55 - 0.71)		0.68 (0.61 - 0.77)		0.73 (0.65 - 0.82)	
1.00	0.38 (0.33 - 0.43)		0.36 (0.30 - 0.43)		0.40 (0.34 - 0.47)		0.47 (0.42 - 0.53)	
2.00	0.20 (0.16 - 0.24)		0.17 (0.14 - 0.22)		0.21 (0.17 - 0.25)		0.21 (0.17 - 0.25)	
5.00	0.06 (0.04 - 0.08)		0.06 (0.04 - 0.09)		0.05 (0.03 - 0.07)		0.07 (0.05 - 0.09)	
10.00	0.02 (0.01 - 0.04)		0.02 (0.01 - 0.03)		0.02 (0.01 - 0.04)		0.02 (0.01 - 0.04)	
20.00	0.00 (0.00 - 0.02)		0.00 (0.00 - 0.02)		0.01 (0.00 - 0.02)		0.00 (0.00 - 0.01)	
50.00	0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)	
Spam Misc%	Ham Misc%							
0.01	12.82 (8.61 - 18.68)		13.11 (8.33 - 20.05)		13.35 (7.63 - 22.32)		13.44 (9.86 - 18.08)	
0.02	9.98 (7.87 - 12.57)		9.26 (6.61 - 12.83)		10.95 (7.83 - 15.10)		10.36 (7.14 - 14.79)	
0.05	6.17 (4.52 - 8.38)		5.59 (4.23 - 7.36)		4.94 (3.59 - 6.77)		5.76 (4.83 - 6.87)	
0.10	3.23 (2.67 - 3.90)		2.86 (2.22 - 3.68)		3.51 (3.03 - 4.05)		3.72 (3.02 - 4.59)	
0.20	1.98 (1.66 - 2.35)		1.79 (1.44 - 2.21)		2.08 (1.76 - 2.46)		2.13 (1.75 - 2.59)	
0.50	0.66 (0.53 - 0.83)		0.70 (0.57 - 0.86)		0.78 (0.64 - 0.95)		0.93 (0.80 - 1.09)	
1.00	0.22 (0.17 - 0.28)		0.23 (0.18 - 0.28)		0.26 (0.21 - 0.32)		0.34 (0.27 - 0.41)	
2.00	0.09 (0.06 - 0.13)		0.08 (0.06 - 0.12)		0.08 (0.06 - 0.11)		0.12 (0.09 - 0.16)	
5.00	0.04 (0.02 - 0.06)		0.03 (0.01 - 0.05)		0.03 (0.01 - 0.05)		0.04 (0.03 - 0.07)	
10.00	0.02 (0.01 - 0.04)		0.01 (0.00 - 0.03)		0.01 (0.01 - 0.03)		0.02 (0.01 - 0.04)	
20.00	0.01 (0.00 - 0.02)		0.00 (0.00 - 0.01)		0.01 (0.00 - 0.02)		0.01 (0.00 - 0.03)	
50.00	0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)	

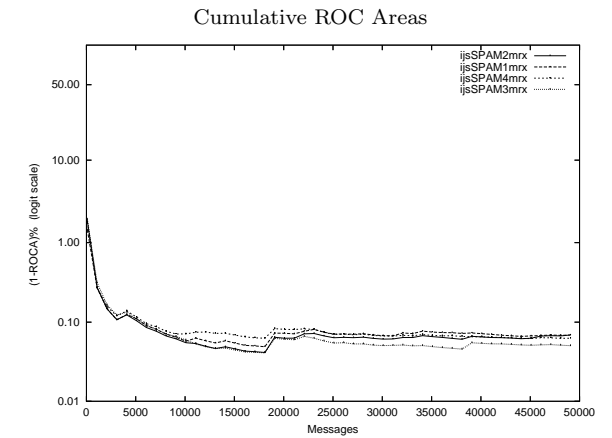
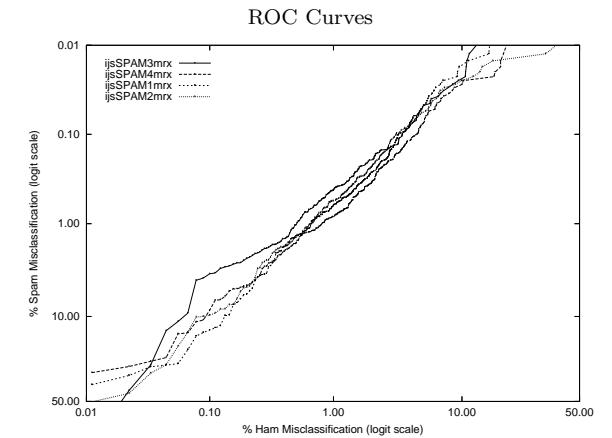


Learning Curves

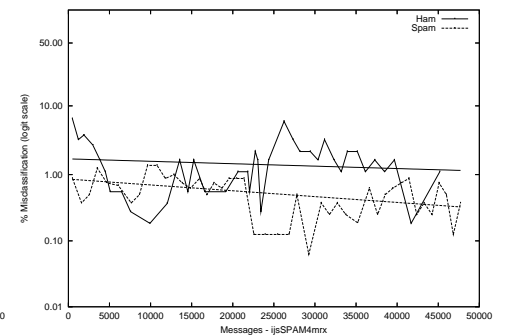
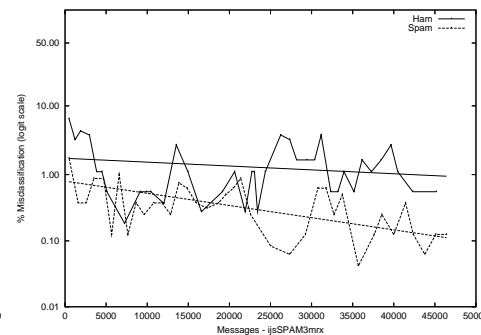
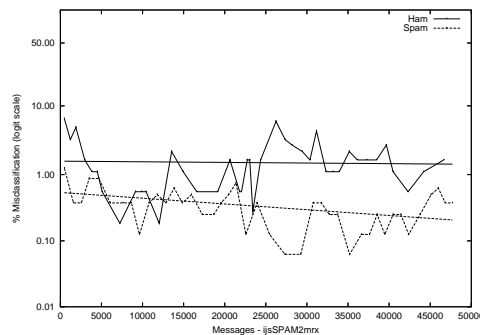
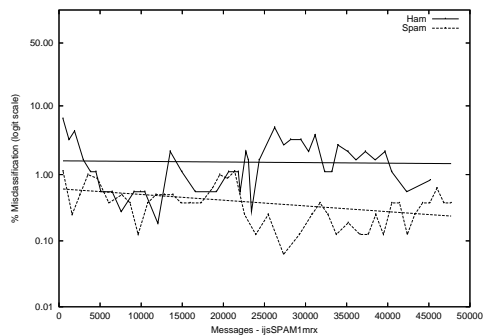


ijs - Jozef Stefan Inst. on Private Mr X. Corpus

Gold Std. Filter Classification	ijsSPAM1mrx		ijsSPAM2mrx		ijsSPAM3mrx		ijsSPAM4mrx	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	1.54 (1.29-1.81)		1.52 (1.27-1.79)		1.33 (1.10-1.59)		1.44 (1.20-1.71)	
Spam Misc%	0.39 (0.33-0.45)		0.34 (0.28-0.40)		0.33 (0.28-0.39)		0.54 (0.47-0.61)	
Logit av Misc%	0.77 (0.69-0.87)		0.72 (0.64-0.81)		0.66 (0.59-0.75)		0.88 (0.79-0.98)	
(1-ROCA)%	0.069 (0.046 - 0.102)		0.069 (0.044 - 0.107)		0.050 (0.028 - 0.091)		0.063 (0.045 - 0.087)	
Ham Misc%	Spam Misc%							
0.01	92.84 (38.97 - 99.62)	94.42 (50.14 - 99.65)	86.21 (47.00 - 97.78)	80.43 (32.56 - 97.22)	92.84 (38.97 - 99.62)	94.42 (50.14 - 99.65)	86.21 (47.00 - 97.78)	80.43 (32.56 - 97.22)
0.02	39.32 (3.67 - 91.67)	50.26 (5.47 - 94.64)	74.51 (17.55 - 97.57)	32.16 (6.25 - 77.12)	39.32 (3.67 - 91.67)	50.26 (5.47 - 94.64)	74.51 (17.55 - 97.57)	32.16 (6.25 - 77.12)
0.05	28.07 (16.12 - 44.22)	27.90 (6.91 - 66.87)	13.80 (1.60 - 61.22)	24.33 (11.44 - 44.45)	28.07 (16.12 - 44.22)	27.90 (6.91 - 66.87)	13.80 (1.60 - 61.22)	24.33 (11.44 - 44.45)
0.10	13.63 (6.48 - 26.44)	9.72 (4.44 - 19.96)	3.56 (1.19 - 10.11)	8.68 (4.04 - 17.67)	13.63 (6.48 - 26.44)	9.72 (4.44 - 19.96)	3.56 (1.19 - 10.11)	8.68 (4.04 - 17.67)
0.20	5.87 (2.95 - 11.34)	4.95 (2.57 - 9.33)	2.42 (1.74 - 3.36)	4.75 (3.12 - 7.16)	5.87 (2.95 - 11.34)	4.95 (2.57 - 9.33)	2.42 (1.74 - 3.36)	4.75 (3.12 - 7.16)
0.50	1.37 (0.98 - 1.91)	1.41 (1.04 - 1.91)	1.03 (0.62 - 1.70)	1.65 (1.17 - 2.32)	1.37 (0.98 - 1.91)	1.41 (1.04 - 1.91)	1.03 (0.62 - 1.70)	1.65 (1.17 - 2.32)
1.00	0.63 (0.49 - 0.81)	0.56 (0.44 - 0.71)	0.42 (0.33 - 0.53)	0.83 (0.67 - 1.03)	0.63 (0.49 - 0.81)	0.56 (0.44 - 0.71)	0.42 (0.33 - 0.53)	0.83 (0.67 - 1.03)
2.00	0.28 (0.21 - 0.36)	0.24 (0.17 - 0.33)	0.19 (0.15 - 0.25)	0.35 (0.26 - 0.49)	0.28 (0.21 - 0.36)	0.24 (0.17 - 0.33)	0.19 (0.15 - 0.25)	0.35 (0.26 - 0.49)
5.00	0.05 (0.03 - 0.09)	0.06 (0.04 - 0.09)	0.05 (0.03 - 0.08)	0.08 (0.05 - 0.13)	0.05 (0.03 - 0.09)	0.06 (0.04 - 0.09)	0.05 (0.03 - 0.08)	0.08 (0.05 - 0.13)
10.00	0.02 (0.01 - 0.04)	0.02 (0.01 - 0.05)	0.02 (0.01 - 0.05)	0.03 (0.01 - 0.06)	0.02 (0.01 - 0.04)	0.02 (0.01 - 0.05)	0.02 (0.01 - 0.05)	0.03 (0.01 - 0.06)
20.00	0.01 (0.00 - 0.03)	0.01 (0.01 - 0.04)	0.00 (0.00 - 0.01)	0.01 (0.00 - 0.04)	0.01 (0.00 - 0.03)	0.01 (0.01 - 0.04)	0.00 (0.00 - 0.01)	0.01 (0.00 - 0.04)
50.00	0.00 (0.00 - 0.01)	0.01 (0.00 - 0.03)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.01)	0.01 (0.00 - 0.03)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)
Spam Misc%	Ham Misc%							
0.01	15.68 (4.30 - 43.49)	39.37 (9.43 - 80.19)	12.69 (9.78 - 16.32)	20.34 (11.70 - 32.98)	15.68 (4.30 - 43.49)	39.37 (9.43 - 80.19)	12.69 (9.78 - 16.32)	20.34 (11.70 - 32.98)
0.02	9.28 (4.73 - 17.41)	13.43 (3.85 - 37.58)	10.62 (7.66 - 14.54)	17.09 (8.38 - 31.73)	9.28 (4.73 - 17.41)	13.43 (3.85 - 37.58)	10.62 (7.66 - 14.54)	17.09 (8.38 - 31.73)
0.05	5.19 (4.05 - 6.63)	5.64 (4.13 - 7.66)	5.05 (3.63 - 6.97)	6.30 (4.67 - 8.43)	5.19 (4.05 - 6.63)	5.64 (4.13 - 7.66)	5.05 (3.63 - 6.97)	6.30 (4.67 - 8.43)
0.10	3.28 (2.62 - 4.09)	3.13 (2.44 - 4.01)	3.53 (2.88 - 4.31)	4.57 (3.65 - 5.71)	3.28 (2.62 - 4.09)	3.13 (2.44 - 4.01)	3.53 (2.88 - 4.31)	4.57 (3.65 - 5.71)
0.20	2.42 (1.99 - 2.94)	2.21 (1.81 - 2.70)	1.94 (1.54 - 2.43)	2.73 (2.14 - 3.48)	2.42 (1.99 - 2.94)	2.21 (1.81 - 2.70)	1.94 (1.54 - 2.43)	2.73 (2.14 - 3.48)
0.50	1.28 (1.04 - 1.58)	1.15 (0.90 - 1.47)	0.87 (0.69 - 1.11)	1.50 (1.22 - 1.86)	1.28 (1.04 - 1.58)	1.15 (0.90 - 1.47)	0.87 (0.69 - 1.11)	1.50 (1.22 - 1.86)
1.00	0.66 (0.51 - 0.86)	0.66 (0.53 - 0.83)	0.51 (0.38 - 0.69)	0.82 (0.64 - 1.04)	0.66 (0.51 - 0.86)	0.66 (0.53 - 0.83)	0.51 (0.38 - 0.69)	0.82 (0.64 - 1.04)
2.00	0.39 (0.28 - 0.54)	0.34 (0.25 - 0.48)	0.27 (0.17 - 0.41)	0.41 (0.30 - 0.56)	0.39 (0.28 - 0.54)	0.34 (0.25 - 0.48)	0.27 (0.17 - 0.41)	0.41 (0.30 - 0.56)
5.00	0.23 (0.15 - 0.35)	0.20 (0.13 - 0.31)	0.08 (0.03 - 0.21)	0.18 (0.11 - 0.29)	0.23 (0.15 - 0.35)	0.20 (0.13 - 0.31)	0.08 (0.03 - 0.21)	0.18 (0.11 - 0.29)
10.00	0.13 (0.08 - 0.23)	0.09 (0.04 - 0.21)	0.07 (0.02 - 0.20)	0.10 (0.05 - 0.18)	0.13 (0.08 - 0.23)	0.09 (0.04 - 0.21)	0.07 (0.02 - 0.20)	0.10 (0.05 - 0.18)
20.00	0.08 (0.03 - 0.18)	0.06 (0.02 - 0.16)	0.04 (0.01 - 0.16)	0.06 (0.02 - 0.13)	0.08 (0.03 - 0.18)	0.06 (0.02 - 0.16)	0.04 (0.01 - 0.16)	0.06 (0.02 - 0.13)
50.00	0.01 (0.00 - 0.04)	0.02 (0.01 - 0.09)	0.02 (0.01 - 0.08)	0.01 (0.00 - 0.04)	0.01 (0.00 - 0.04)	0.02 (0.01 - 0.09)	0.02 (0.01 - 0.08)	0.01 (0.00 - 0.04)

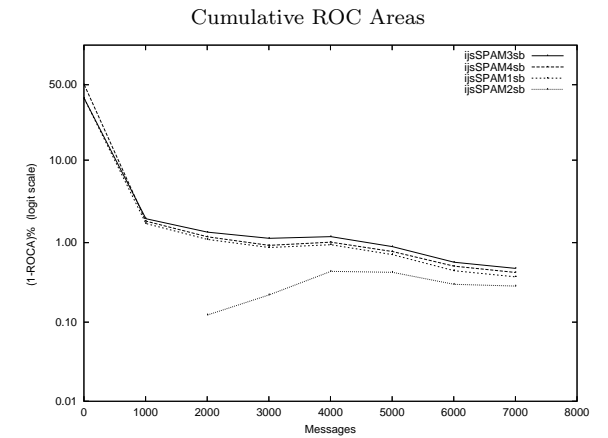
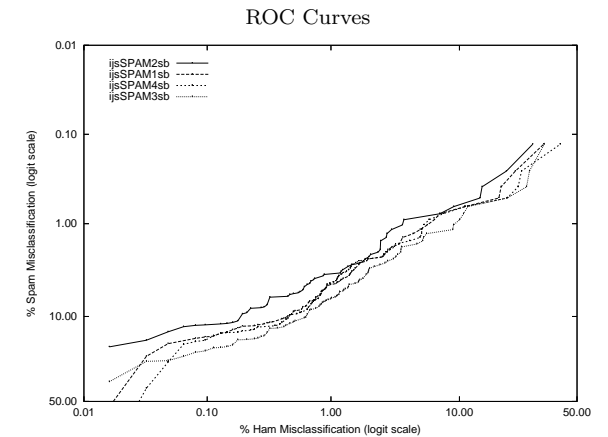


Learning Curves

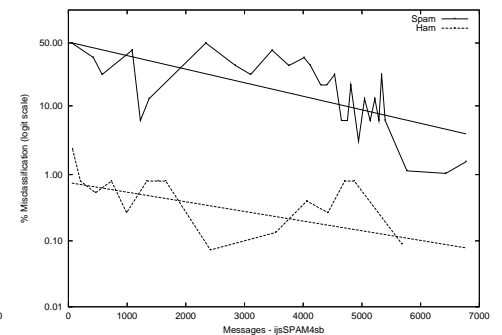
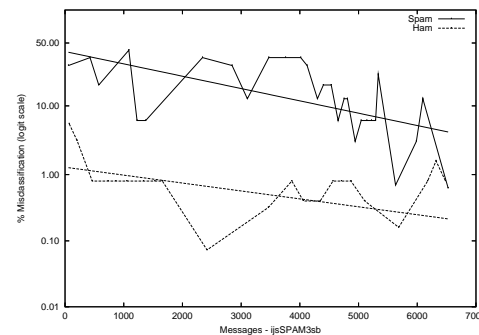
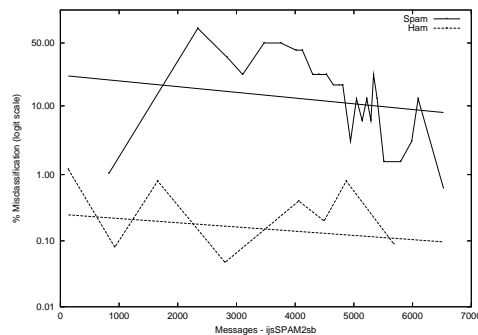
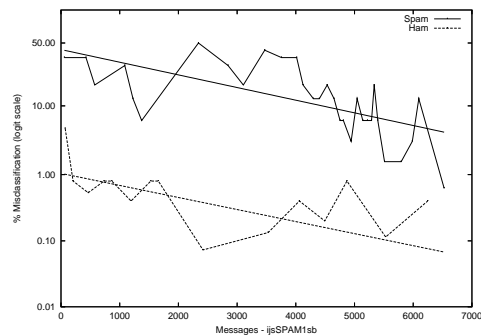


ijs - Jozef Stefan Inst. on Private S.B. Corpus

Gold Std. Filter Classification	ijsSPAM1sb			ijsSPAM2sb			ijsSPAM3sb			ijsSPAM4sb		
	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	
Ham Misc%	0.35 (0.22-0.53)			0.16 (0.08-0.29)			0.59 (0.42-0.82)			0.30 (0.18-0.48)		
Spam Misc%	11.35 (9.21-13.80)			11.74 (9.56-14.22)			11.10 (8.97-13.52)			12.77 (10.50-15.33)		
Logit av Misc%	2.09 (1.65-2.63)			1.44 (1.04-1.99)			2.66 (2.19-3.22)			2.07 (1.62-2.64)		
(1-ROCA)%	0.372 (0.220 - 0.628)			0.285 (0.159 - 0.508)			0.475 (0.301 - 0.749)			0.422 (0.242 - 0.735)		
Ham Misc%	Spam Misc%											
0.01	77.42 (34.49 - 95.71)			71.48 (19.24 - 96.35)			81.29 (33.57 - 97.39)			100.00 (nan - nan)		
0.02	53.55 (12.35 - 90.41)			19.61 (2.89 - 66.65)			37.55 (6.94 - 82.90)			80.00 (nan - nan)		
0.05	18.32 (6.79 - 40.84)			14.19 (5.64 - 31.40)			25.81 (11.41 - 48.43)			26.58 (6.27 - 66.22)		
0.10	15.87 (11.26 - 21.90)			12.13 (9.13 - 15.94)			21.29 (15.42 - 28.63)			17.03 (10.81 - 25.80)		
0.20	12.52 (9.63 - 16.12)			9.42 (6.48 - 13.50)			16.90 (13.14 - 21.48)			13.94 (11.23 - 17.16)		
0.50	8.90 (6.65 - 11.82)			5.81 (4.17 - 8.04)			11.35 (8.58 - 14.88)			9.55 (6.98 - 12.93)		
1.00	4.65 (3.23 - 6.64)			3.61 (2.43 - 5.33)			6.58 (4.87 - 8.84)			4.52 (2.86 - 7.06)		
2.00	2.45 (1.47 - 4.07)			2.45 (1.37 - 4.36)			3.35 (2.09 - 5.34)			2.45 (1.50 - 3.99)		
5.00	1.29 (0.64 - 2.57)			0.90 (0.38 - 2.11)			1.68 (0.93 - 3.02)			1.55 (0.81 - 2.94)		
10.00	0.77 (0.30 - 2.01)			0.65 (0.20 - 2.03)			1.03 (0.42 - 2.53)			0.77 (0.31 - 1.95)		
20.00	0.39 (0.11 - 1.34)			0.39 (0.11 - 1.38)			0.65 (0.22 - 1.88)			0.65 (0.21 - 1.93)		
50.00	0.06 (0.06 - 0.07)			0.06 (0.06 - 0.07)			0.13 (0.04 - 0.41)			0.13 (0.03 - 0.49)		
Spam Misc%	Ham Misc%											
0.01	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.10	47.65 (28.09 - 67.96)			45.74 (23.04 - 70.36)			54.74 (31.74 - 75.89)			56.68 (33.62 - 77.16)		
0.20	34.87 (14.21 - 63.38)			30.35 (11.26 - 59.95)			35.13 (15.05 - 62.34)			42.23 (15.20 - 74.89)		
0.50	19.48 (6.05 - 47.64)			14.51 (4.30 - 39.08)			28.02 (11.24 - 54.49)			24.94 (6.26 - 62.32)		
1.00	6.50 (2.04 - 18.82)			3.79 (1.10 - 12.19)			10.03 (4.07 - 22.65)			5.94 (2.03 - 16.14)		
2.00	2.84 (1.43 - 5.56)			2.47 (1.58 - 3.85)			3.59 (1.89 - 6.75)			2.95 (1.33 - 6.43)		
5.00	0.93 (0.60 - 1.44)			0.61 (0.35 - 1.06)			1.38 (0.91 - 2.09)			0.93 (0.66 - 1.32)		
10.00	0.45 (0.25 - 0.81)			0.19 (0.10 - 0.37)			0.64 (0.43 - 0.96)			0.47 (0.28 - 0.77)		
20.00	0.05 (0.01 - 0.19)			0.02 (0.00 - 0.07)			0.11 (0.04 - 0.36)			0.06 (0.02 - 0.22)		
50.00	0.03 (0.01 - 0.12)			0.02 (0.00 - 0.05)			0.02 (0.00 - 0.06)			0.03 (0.01 - 0.11)		

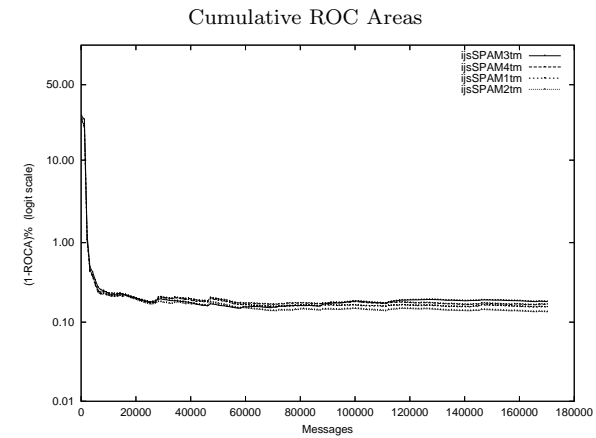
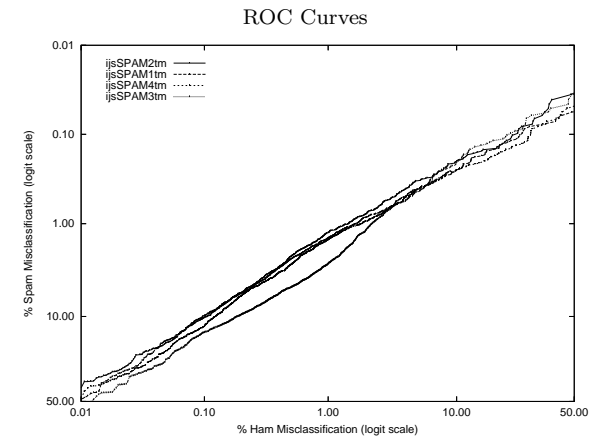


Learning Curves

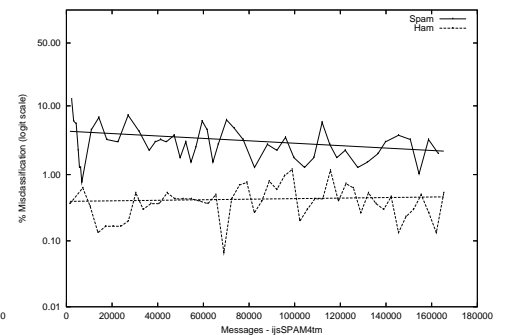
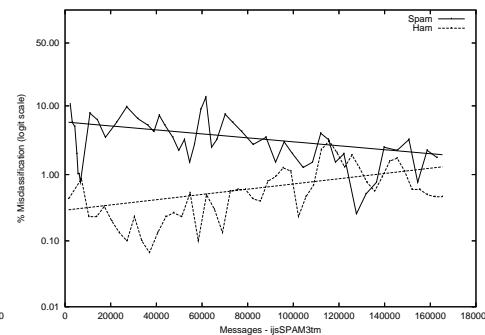
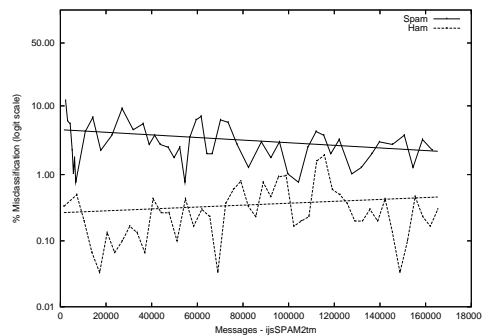
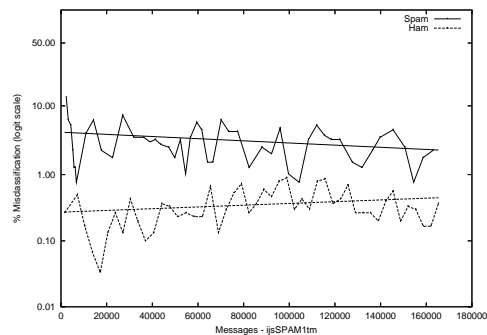


ijs - Jozef Stefan Inst. on Private T.M. Corpus

Gold Std. Filter Classification	ijsSPAM1tm		ijsSPAM2tm		ijsSPAM3tm		ijsSPAM4tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.36 (0.33-0.39)		0.36 (0.33-0.39)		0.70 (0.66-0.75)		0.43 (0.40-0.46)	
Spam Misc%	3.31 (3.06-3.57)		3.43 (3.18-3.70)		3.87 (3.61-4.15)		3.34 (3.09-3.60)	
Logit av Misc%	1.10 (1.03-1.16)		1.12 (1.06-1.18)		1.66 (1.59-1.74)		1.21 (1.14-1.27)	
(1-ROCA)%	0.155 (0.130 - 0.184)		0.135 (0.111 - 0.163)		0.181 (0.155 - 0.212)		0.167 (0.139 - 0.201)	
Ham Misc%	Spam Misc%							
0.01	48.52 (35.82 - 61.42)		41.60 (34.13 - 49.47)		51.54 (44.09 - 58.92)		48.39 (38.15 - 58.76)	
0.02	32.77 (27.06 - 39.05)		31.21 (25.73 - 37.27)		39.66 (32.66 - 47.11)		34.60 (30.47 - 38.98)	
0.05	18.01 (15.29 - 21.08)		17.07 (14.67 - 19.78)		25.29 (20.77 - 30.40)		21.33 (17.53 - 25.69)	
0.10	9.88 (8.61 - 11.32)		10.31 (9.16 - 11.58)		14.49 (13.18 - 15.90)		12.66 (10.75 - 14.85)	
0.20	5.69 (5.04 - 6.42)		5.73 (5.24 - 6.26)		9.52 (8.71 - 10.40)		6.12 (5.38 - 6.96)	
0.50	2.49 (2.23 - 2.79)		2.32 (2.07 - 2.60)		5.02 (4.67 - 5.40)		2.85 (2.52 - 3.23)	
1.00	1.44 (1.27 - 1.63)		1.26 (1.11 - 1.43)		2.78 (2.52 - 3.06)		1.52 (1.32 - 1.74)	
2.00	0.86 (0.74 - 0.99)		0.78 (0.66 - 0.93)		1.19 (1.05 - 1.35)		0.94 (0.81 - 1.10)	
5.00	0.42 (0.34 - 0.53)		0.33 (0.27 - 0.42)		0.45 (0.36 - 0.55)		0.42 (0.34 - 0.52)	
10.00	0.25 (0.19 - 0.33)		0.20 (0.16 - 0.27)		0.21 (0.15 - 0.29)		0.25 (0.18 - 0.34)	
20.00	0.13 (0.09 - 0.19)		0.13 (0.09 - 0.19)		0.12 (0.08 - 0.18)		0.16 (0.11 - 0.24)	
50.00	0.06 (0.03 - 0.11)		0.04 (0.01 - 0.09)		0.04 (0.01 - 0.09)		0.05 (0.03 - 0.10)	
Spam Misc%	Ham Misc%							
0.01	98.66 (81.80 - 99.92)		98.90 (74.45 - 99.96)		99.54 (76.97 - 99.99)		98.63 (80.95 - 99.92)	
0.02	95.30 (56.47 - 99.69)		74.13 (5.64 - 99.28)		86.01 (24.37 - 99.15)		97.16 (57.57 - 99.88)	
0.05	58.25 (38.63 - 75.57)		38.48 (15.30 - 68.42)		47.10 (18.34 - 77.93)		52.37 (32.51 - 71.50)	
0.10	27.43 (16.98 - 41.13)		25.77 (17.23 - 36.67)		23.29 (13.90 - 36.35)		30.69 (22.13 - 40.82)	
0.20	12.63 (8.91 - 17.61)		10.01 (6.53 - 15.02)		10.22 (7.58 - 13.66)		15.70 (9.01 - 25.94)	
0.50	4.09 (3.30 - 5.07)		3.41 (2.76 - 4.22)		4.44 (3.66 - 5.38)		4.33 (3.58 - 5.22)	
1.00	1.69 (1.38 - 2.07)		1.46 (1.21 - 1.77)		2.24 (2.01 - 2.49)		1.87 (1.45 - 2.42)	
2.00	0.66 (0.57 - 0.75)		0.58 (0.50 - 0.66)		1.35 (1.25 - 1.46)		0.72 (0.62 - 0.84)	
5.00	0.23 (0.20 - 0.27)		0.24 (0.22 - 0.27)		0.50 (0.45 - 0.57)		0.25 (0.22 - 0.29)	
10.00	0.10 (0.08 - 0.12)		0.11 (0.09 - 0.12)		0.18 (0.16 - 0.21)		0.12 (0.11 - 0.14)	
20.00	0.04 (0.03 - 0.05)		0.04 (0.03 - 0.05)		0.07 (0.06 - 0.08)		0.05 (0.04 - 0.07)	
50.00	0.01 (0.01 - 0.02)		0.00 (0.00 - 0.01)		0.01 (0.01 - 0.02)		0.01 (0.01 - 0.02)	



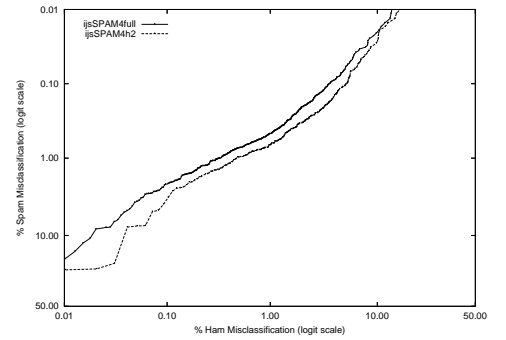
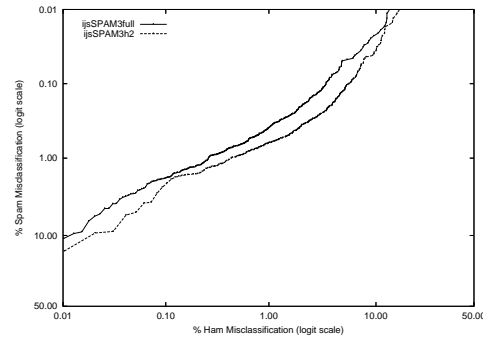
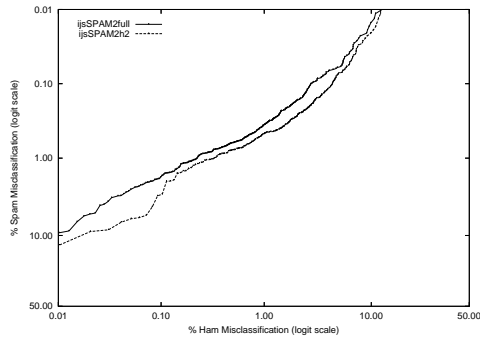
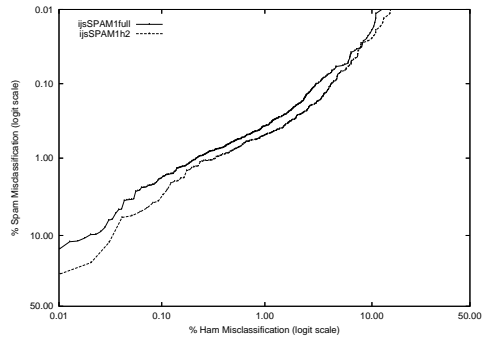
Learning Curves



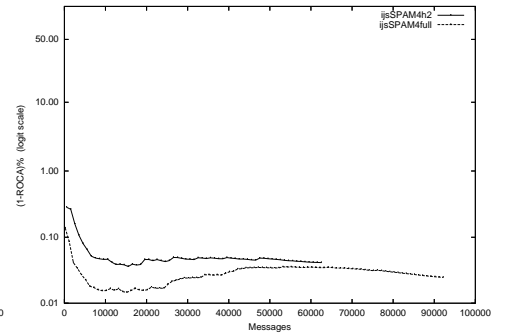
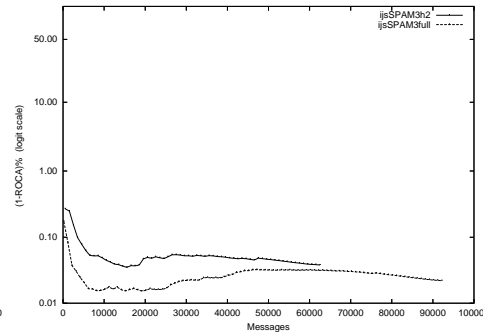
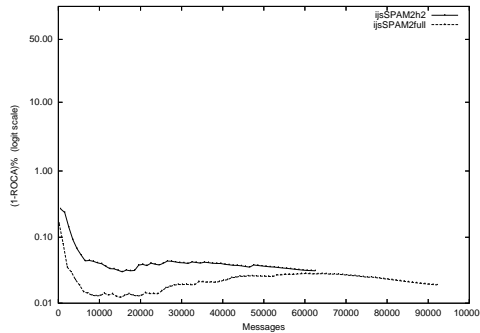
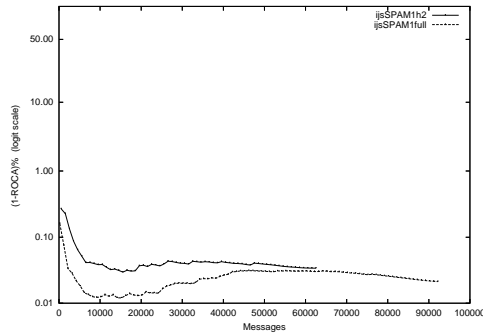
ijs - Jozef Stefan Inst. - Public subset Corpora

	ijsSPAM1full	ijsSPAM1h2
Ham Misc%	0.25 (0.20-0.30)	0.32 (0.22-0.45)
Spam Misc%	0.93 (0.85-1.02)	1.02 (0.94-1.11)
Logit av Misc%	0.48 (0.43-0.53)	0.57 (0.48-0.68)
(1-ROCA)%	0.021 (0.018 - 0.025)	0.034 (0.024 - 0.048)
	ijsSPAM2full	ijsSPAM2h2
Ham Misc%	0.23 (0.18-0.28)	0.30 (0.20-0.43)
Spam Misc%	0.95 (0.87-1.04)	1.04 (0.96-1.13)
Logit av Misc%	0.47 (0.42-0.52)	0.56 (0.46-0.67)
(1-ROCA)%	0.019 (0.015 - 0.023)	0.031 (0.023 - 0.042)
	ijsSPAM3full	ijsSPAM3h2
Ham Misc%	0.26 (0.21-0.32)	0.38 (0.27-0.52)
Spam Misc%	0.97 (0.89-1.06)	1.11 (1.02-1.20)
Logit av Misc%	0.51 (0.45-0.56)	0.65 (0.55-0.77)
(1-ROCA)%	0.022 (0.019 - 0.026)	0.038 (0.028 - 0.051)
	ijsSPAM4full	ijsSPAM4h2
Ham Misc%	0.37 (0.31-0.43)	0.45 (0.33-0.61)
Spam Misc%	0.91 (0.83-0.99)	1.05 (0.97-1.14)
Logit av Misc%	0.58 (0.53-0.63)	0.69 (0.59-0.80)
(1-ROCA)%	0.025 (0.021 - 0.029)	0.041 (0.030 - 0.056)

ROC Curves

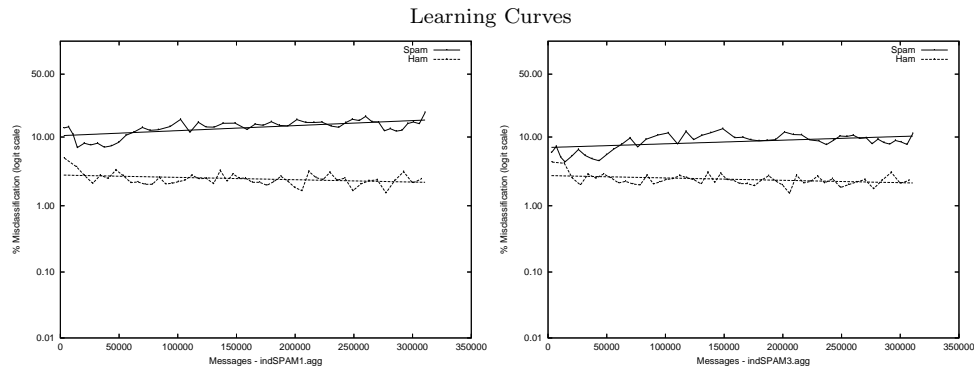
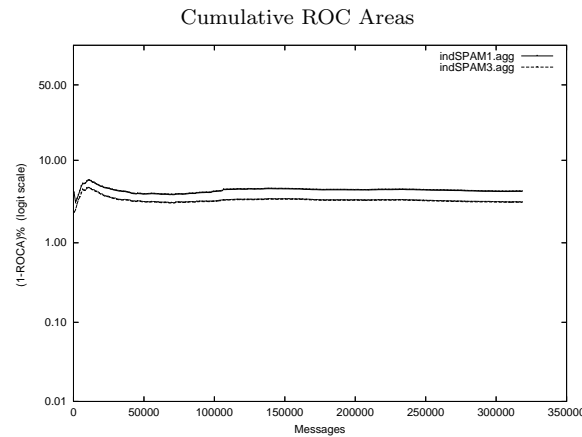
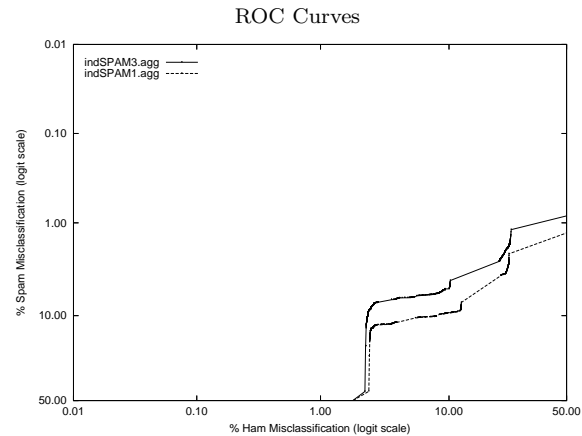


Cumulative ROC Area



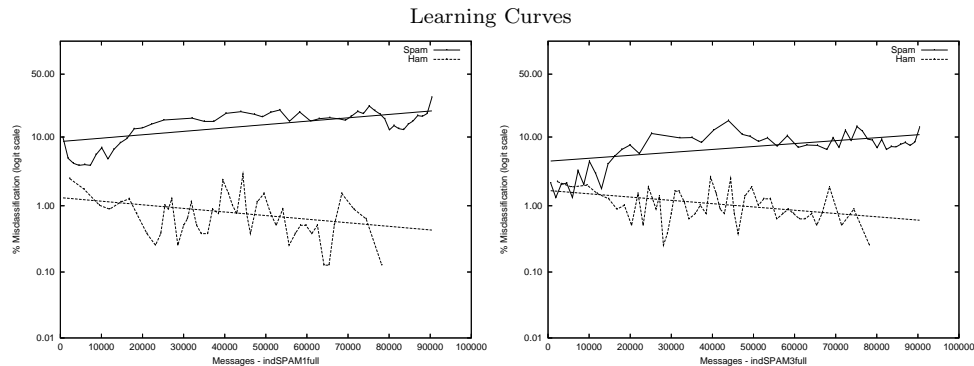
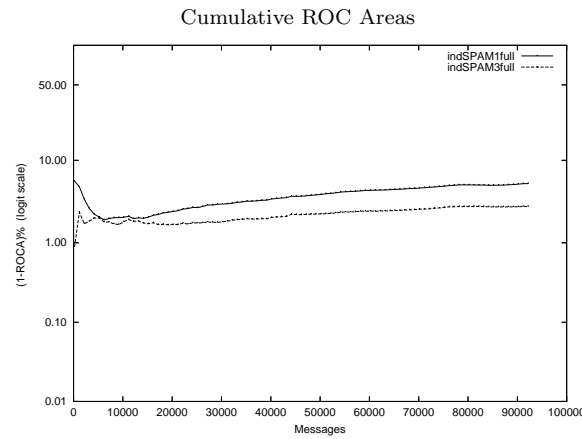
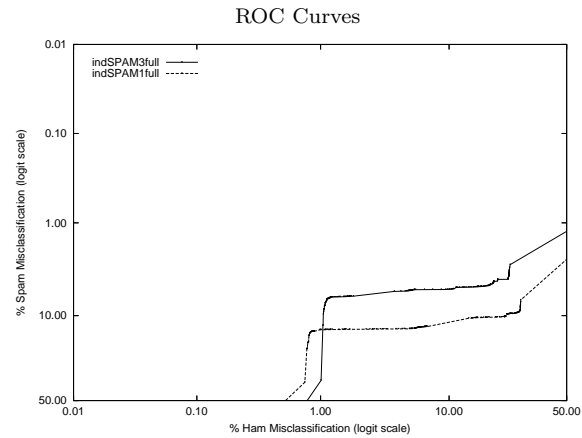
ind - Indiana U. (Yang) on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	indSPAM1.agg		indSPAM3.agg	
	ham spam	ham spam	ham spam	ham spam
Ham Misc%	2.54 (2.47-2.60)		2.49 (2.42-2.55)	
Spam Misc%	13.57 (13.37-13.77)		8.74 (8.57-8.90)	
Logit av Misc%	6.01 (5.92-6.10)		4.71 (4.63-4.79)	
(1-ROCA)%	4.302 (4.229 - 4.377)		3.168 (3.094 - 3.244)	
Ham Misc%	Spam Misc%			
0.01	97.90 (97.58 - 98.18)	98.76 (98.59 - 98.92)	98.76 (98.59 - 98.92)	98.76 (98.59 - 98.92)
0.02	97.67 (97.49 - 97.83)	98.61 (98.42 - 98.78)	98.61 (98.42 - 98.78)	98.61 (98.42 - 98.78)
0.05	97.10 (96.88 - 97.32)	98.03 (97.74 - 98.28)	98.03 (97.74 - 98.28)	98.03 (97.74 - 98.28)
0.10	96.06 (95.75 - 96.34)	96.99 (96.62 - 97.32)	96.99 (96.62 - 97.32)	96.99 (96.62 - 97.32)
0.20	93.78 (93.30 - 94.23)	94.53 (94.01 - 95.01)	94.53 (94.01 - 95.01)	94.53 (94.01 - 95.01)
0.50	87.16 (86.43 - 87.86)	87.17 (86.46 - 87.84)	87.17 (86.46 - 87.84)	87.17 (86.46 - 87.84)
1.00	76.15 (75.17 - 77.10)	75.15 (73.92 - 76.34)	75.15 (73.92 - 76.34)	75.15 (73.92 - 76.34)
2.00	54.05 (52.68 - 55.42)	50.89 (49.21 - 52.57)	50.89 (49.21 - 52.57)	50.89 (49.21 - 52.57)
5.00	10.99 (10.81 - 11.17)	6.54 (6.39 - 6.68)	6.54 (6.39 - 6.68)	6.54 (6.39 - 6.68)
10.00	9.41 (9.27 - 9.56)	5.26 (5.00 - 5.55)	5.26 (5.00 - 5.55)	5.26 (5.00 - 5.55)
20.00	4.75 (4.62 - 4.87)	2.95 (2.84 - 3.06)	2.95 (2.84 - 3.06)	2.95 (2.84 - 3.06)
50.00	1.51 (1.45 - 1.58)	0.90 (0.85 - 0.96)	0.90 (0.85 - 0.96)	0.90 (0.85 - 0.96)
Spam Misc%	Ham Misc%			
0.01	99.16 (98.99 - 99.30)	99.17 (99.00 - 99.31)	99.17 (99.00 - 99.31)	99.17 (99.00 - 99.31)
0.02	98.92 (98.70 - 99.11)	98.88 (98.48 - 99.17)	98.88 (98.48 - 99.17)	98.88 (98.48 - 99.17)
0.05	98.43 (98.22 - 98.61)	98.35 (98.09 - 98.58)	98.35 (98.09 - 98.58)	98.35 (98.09 - 98.58)
0.10	97.51 (96.42 - 98.28)	97.63 (97.02 - 98.12)	97.63 (97.02 - 98.12)	97.63 (97.02 - 98.12)
0.20	93.62 (93.04 - 94.15)	95.15 (94.52 - 95.71)	95.15 (94.52 - 95.71)	95.15 (94.52 - 95.71)
0.50	88.87 (85.66 - 91.44)	81.35 (77.68 - 84.54)	81.35 (77.68 - 84.54)	81.35 (77.68 - 84.54)
1.00	68.64 (66.58 - 70.63)	42.96 (38.06 - 47.99)	42.96 (38.06 - 47.99)	42.96 (38.06 - 47.99)
2.00	32.59 (29.81 - 35.51)	24.27 (23.81 - 24.74)	24.27 (23.81 - 24.74)	24.27 (23.81 - 24.74)
5.00	19.22 (18.85 - 19.60)	10.11 (9.97 - 10.26)	10.11 (9.97 - 10.26)	10.11 (9.97 - 10.26)
10.00	8.20 (7.99 - 8.42)	2.38 (2.31 - 2.46)	2.38 (2.31 - 2.46)	2.38 (2.31 - 2.46)
20.00	2.47 (2.40 - 2.53)	2.29 (2.22 - 2.36)	2.29 (2.22 - 2.36)	2.29 (2.22 - 2.36)
50.00	2.18 (2.11 - 2.24)	2.03 (1.96 - 2.10)	2.03 (1.96 - 2.10)	2.03 (1.96 - 2.10)



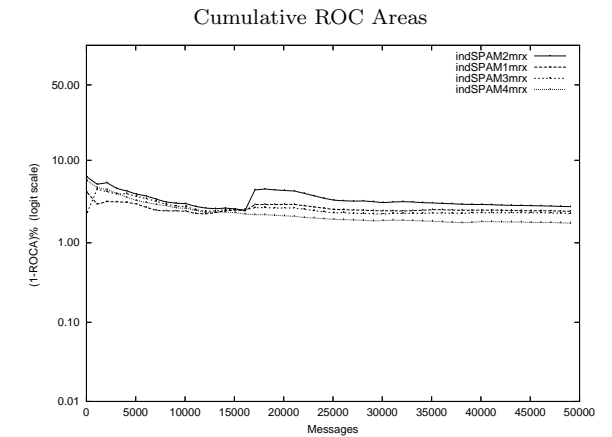
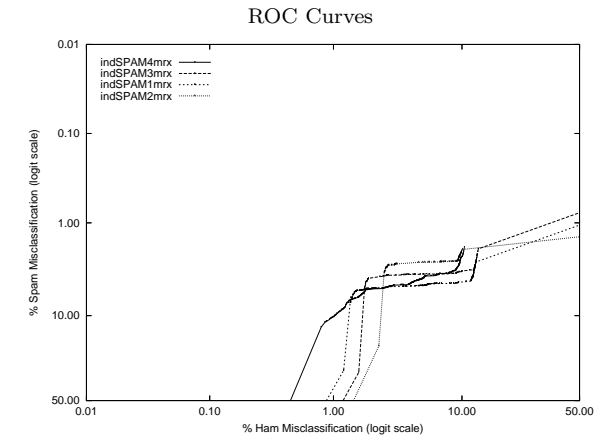
ind - Indiana U. (Yang) on Public trec05p-1/full Corpus

Gold Std.	indSPAM1full		indSPAM3full			
	ham	spam	ham	spam		
Filter	ham	39077	8003	ham	38971	4043
Classification	spam	322	44787	spam	428	48747
Ham Misc%	0.82 (0.73-0.91)		1.09 (0.99-1.19)			
Spam Misc%	15.16 (14.86-15.47)		7.66 (7.43-7.89)			
Logit av Misc%	3.70 (3.50-3.90)		2.93 (2.79-3.08)			
(1-ROCA)%	5.346 (5.205 - 5.491)		2.822 (2.710 - 2.940)			
Ham Misc%	Spam Misc%					
0.01	97.62 (95.18 - 98.84)		98.48 (98.04 - 98.83)			
0.02	96.62 (95.99 - 97.15)		98.12 (97.61 - 98.52)			
0.05	95.58 (94.33 - 96.57)		97.77 (97.55 - 97.98)			
0.10	93.19 (90.63 - 95.09)		97.35 (97.12 - 97.56)			
0.20	85.15 (81.59 - 88.13)		93.84 (90.84 - 95.90)			
0.50	60.75 (54.21 - 66.93)		72.38 (67.30 - 76.95)			
1.00	13.82 (13.48 - 14.18)		37.77 (10.77 - 75.32)			
2.00	13.65 (13.35 - 13.97)		6.24 (6.01 - 6.47)			
5.00	13.51 (13.21 - 13.82)		5.52 (5.32 - 5.73)			
10.00	11.84 (11.55 - 12.13)		5.33 (5.14 - 5.54)			
20.00	10.38 (10.12 - 10.65)		4.70 (4.46 - 4.95)			
50.00	3.73 (3.57 - 3.90)		1.69 (1.58 - 1.81)			
Spam Misc%	Ham Misc%					
0.01	98.47 (96.69 - 99.30)		98.37 (97.74 - 98.83)			
0.02	97.28 (95.88 - 98.21)		97.31 (96.15 - 98.13)			
0.05	95.32 (93.90 - 96.43)		94.83 (93.26 - 96.05)			
0.10	87.82 (77.95 - 93.63)		86.32 (78.89 - 91.42)			
0.20	73.46 (72.11 - 74.78)		76.83 (76.32 - 77.33)			
0.50	70.37 (69.54 - 71.18)		72.70 (72.23 - 73.17)			
1.00	68.46 (67.80 - 69.12)		63.78 (61.94 - 65.58)			
2.00	61.76 (60.86 - 62.64)		43.84 (41.21 - 46.51)			
5.00	42.48 (41.26 - 43.72)		16.07 (11.06 - 22.76)			
10.00	24.44 (23.86 - 25.02)		1.05 (0.95 - 1.16)			
20.00	0.78 (0.69 - 0.88)		1.03 (0.93 - 1.14)			
50.00	0.61 (0.53 - 0.70)		0.81 (0.73 - 0.90)			

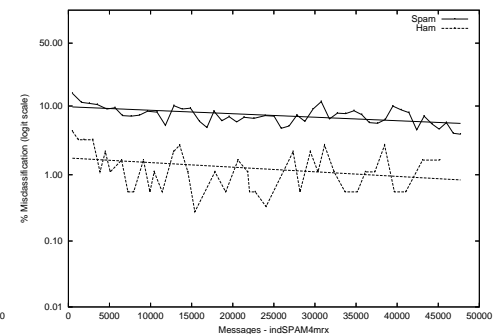
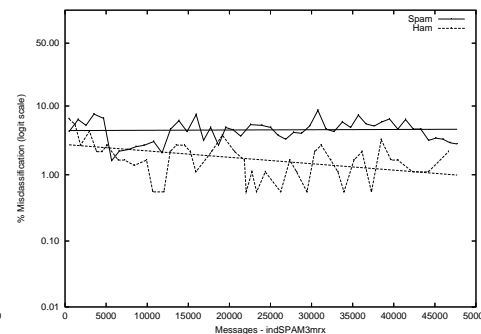
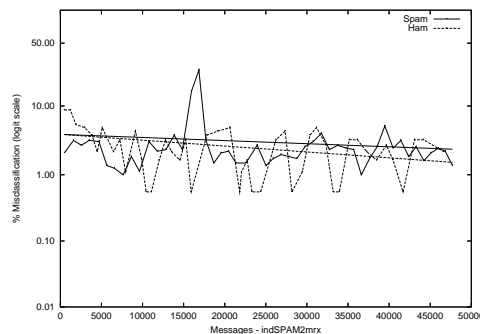
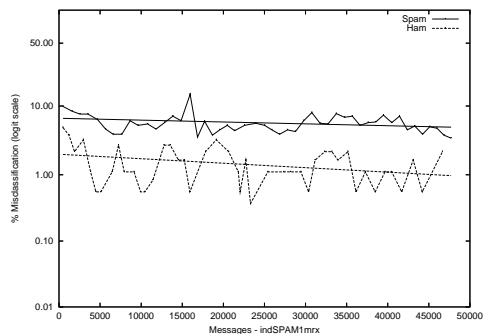


ind - Indiana U. (Yang) on Private Mr X. Corpus

Gold Std. Filter Classification	indSPAM1mrx		indSPAM2mrx		indSPAM3mrx		indSPAM4mrx	
	ham	spam	ham	spam	ham	spam	ham	spam
8905	2334	8798	1237	8874	1849	8922	2999	
133	37714	240	38811	164	38199	116	37049	
Ham Misc%	1.47 (1.23-1.74)		2.66 (2.33-3.01)		1.81 (1.55-2.11)		1.28 (1.06-1.54)	
Spam Misc%	5.83 (5.60-6.06)		3.09 (2.92-3.26)		4.62 (4.41-4.83)		7.49 (7.23-7.75)	
Logit av Misc%	2.95 (2.71-3.21)		2.86 (2.68-3.07)		2.90 (2.69-3.14)		3.14 (2.87-3.44)	
(1-ROCA)%	2.471 (2.294 - 2.662)		2.804 (2.577 - 3.049)		2.321 (2.133 - 2.526)		1.757 (1.611 - 1.915)	
Ham Misc%	Spam Misc%							
0.01	99.89 (99.42 - 99.98)		100.00 (nan - nan)		99.73 (99.58 - 99.83)		99.94 (nan - nan)	
0.02	99.46 (97.54 - 99.88)		99.97 (nan - nan)		99.68 (99.50 - 99.80)		99.64 (98.02 - 99.93)	
0.05	99.15 (98.81 - 99.39)		99.76 (99.27 - 99.92)		99.59 (99.35 - 99.74)		98.52 (95.86 - 99.48)	
0.10	99.10 (97.42 - 99.69)		99.75 (99.66 - 99.81)		99.31 (98.73 - 99.63)		97.33 (94.67 - 98.68)	
0.20	91.37 (75.98 - 97.26)		99.71 (97.67 - 99.97)		95.83 (83.82 - 99.03)		94.26 (84.90 - 97.96)	
0.50	78.65 (68.90 - 85.97)		88.25 (80.45 - 93.20)		82.71 (74.81 - 88.51)		43.47 (25.71 - 63.07)	
1.00	45.96 (33.28 - 59.18)		68.82 (59.98 - 76.48)		59.55 (50.00 - 68.43)		10.26 (8.18 - 12.80)	
2.00	5.25 (5.01 - 5.51)		30.64 (18.06 - 46.96)		4.04 (2.74 - 5.92)		5.30 (4.90 - 5.74)	
5.00	4.88 (4.65 - 5.12)		2.72 (2.57 - 2.88)		3.67 (3.50 - 3.85)		4.16 (3.87 - 4.46)	
10.00	4.48 (4.27 - 4.69)		2.03 (1.71 - 2.39)		3.44 (3.26 - 3.63)		2.36 (1.81 - 3.07)	
20.00	2.45 (2.31 - 2.60)		1.86 (1.74 - 1.98)		1.76 (1.63 - 1.90)		1.66 (1.55 - 1.78)	
50.00	1.41 (1.31 - 1.52)		1.50 (1.39 - 1.61)		1.05 (0.96 - 1.15)		1.07 (0.98 - 1.18)	
Spam Misc%	Ham Misc%							
0.01	99.14 (98.36 - 99.55)		99.88 (99.67 - 99.95)		99.09 (98.37 - 99.50)		99.48 (nan - nan)	
0.02	98.16 (95.58 - 99.25)		99.67 (98.86 - 99.90)		98.16 (95.68 - 99.23)		99.29 (97.96 - 99.76)	
0.05	95.36 (93.45 - 96.74)		98.43 (96.67 - 99.27)		95.42 (93.59 - 96.75)		98.16 (96.97 - 98.89)	
0.10	93.80 (93.31 - 94.26)		97.27 (96.87 - 97.62)		93.82 (92.51 - 94.91)		96.33 (94.66 - 97.49)	
0.20	93.80 (93.31 - 94.26)		88.02 (85.83 - 89.90)		82.20 (81.05 - 83.29)		92.01 (89.82 - 93.76)	
0.50	74.22 (73.24 - 75.17)		85.56 (84.70 - 86.39)		75.13 (71.99 - 78.02)		79.14 (75.57 - 82.32)	
1.00	63.19 (60.53 - 65.77)		79.96 (77.67 - 82.08)		52.00 (47.94 - 56.04)		53.23 (48.34 - 58.06)	
2.00	32.42 (28.55 - 36.54)		10.05 (7.25 - 13.76)		13.07 (11.48 - 14.83)		10.32 (9.57 - 11.13)	
5.00	2.81 (1.51 - 5.17)		2.52 (2.20 - 2.90)		1.79 (1.54 - 2.09)		2.77 (2.31 - 3.31)	
10.00	1.35 (1.13 - 1.61)		2.43 (2.11 - 2.81)		1.75 (1.50 - 2.04)		1.03 (0.82 - 1.28)	
20.00	1.25 (1.04 - 1.50)		2.30 (1.99 - 2.66)		1.65 (1.41 - 1.93)		0.75 (0.58 - 0.97)	
50.00	0.98 (0.81 - 1.20)		1.37 (1.14 - 1.65)		1.18 (0.99 - 1.42)		0.45 (0.33 - 0.62)	

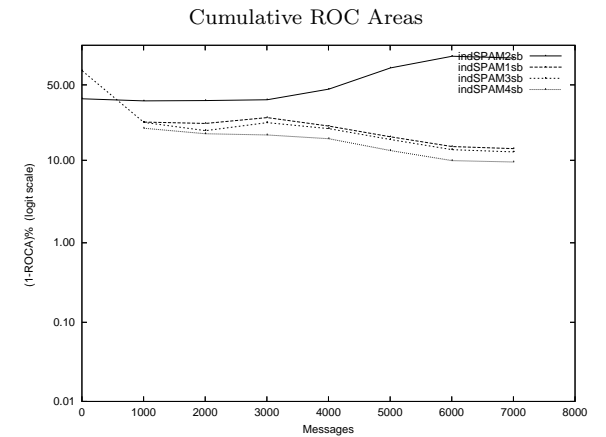
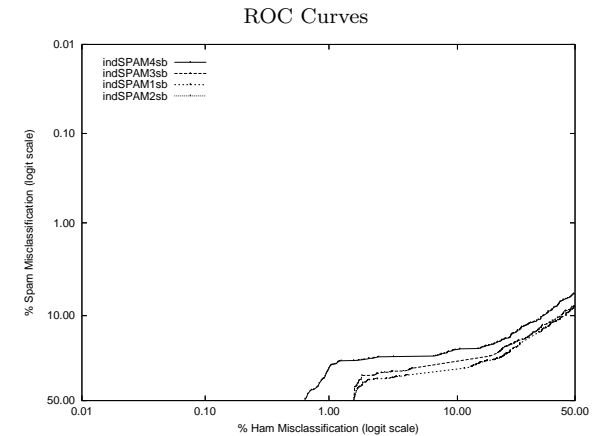


Learning Curves

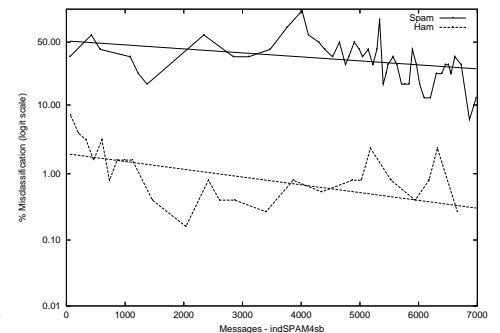
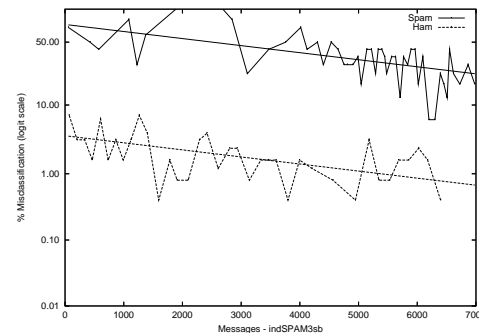
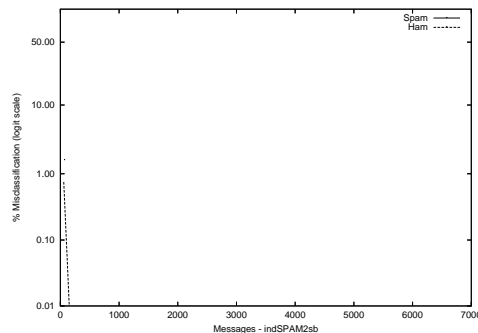
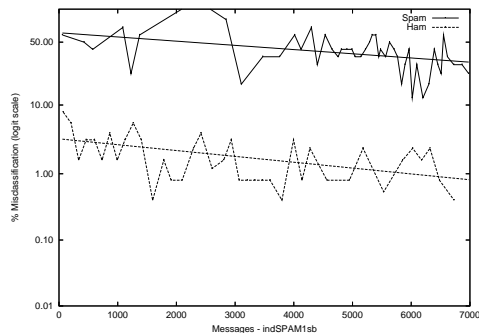


ind - Indiana U. (Yang) on Private S.B. Corpus

Gold Std. Filter Classification	indSPAM1sb		indSPAM2sb		indSPAM3sb		indSPAM4sb	
	ham	spam	ham	spam	ham	spam	ham	spam
ham	6117	318	6229	775	6117	287	6173	273
spam	114	457	2	0	114	488	58	502
Ham Misc%	1.83 (1.51-2.19)		0.03 (0.00-0.12)		1.83 (1.51-2.19)		0.93 (0.71-1.20)	
Spam Misc%	41.03 (37.54-44.59)		100.00 (99.61-100.00)		37.03 (33.62-40.54)		35.23 (31.86-38.70)	
Logit av Misc%	10.22 (9.20-11.35)		99.99 (0.00-nan)		9.48 (8.51-10.54)		6.67 (5.80-7.66)	
(1-ROCA)%	13.507 (12.047 - 15.112)		68.572 (66.684 - 70.399)		12.454 (11.002 - 14.068)		9.588 (8.095 - 11.324)	
Ham Misc%	Spam Misc%							
0.01	98.58 (97.15 - 99.30)		100.00 (nan - nan)		97.68 (94.77 - 98.99)		100.00 (nan - nan)	
0.02	98.19 (96.20 - 99.15)		100.00 (nan - nan)		96.77 (91.86 - 98.76)		100.00 (nan - nan)	
0.05	97.81 (94.17 - 99.19)		100.00 (nan - nan)		92.65 (86.84 - 96.01)		100.00 (nan - nan)	
0.10	93.16 (84.08 - 97.23)		99.87 (nan - nan)		91.10 (86.71 - 94.13)		100.00 (nan - nan)	
0.20	91.48 (89.32 - 93.24)		99.87 (nan - nan)		88.77 (85.90 - 91.12)		99.35 (nan - nan)	
0.50	89.03 (86.22 - 91.33)		99.87 (nan - nan)		86.19 (83.00 - 88.87)		63.61 (27.68 - 88.87)	
1.00	83.35 (79.20 - 86.81)		99.35 (97.88 - 99.81)		80.65 (76.93 - 83.89)		30.32 (22.24 - 39.84)	
2.00	38.06 (30.96 - 45.72)		98.58 (97.33 - 99.25)		34.32 (30.38 - 38.50)		25.68 (22.58 - 29.05)	
5.00	33.42 (30.26 - 36.74)		97.16 (95.44 - 98.24)		30.06 (26.86 - 33.47)		24.26 (21.37 - 27.40)	
10.00	30.58 (27.38 - 33.98)		94.45 (92.46 - 95.94)		27.61 (24.45 - 31.02)		21.16 (18.51 - 24.07)	
20.00	24.65 (21.73 - 27.81)		88.52 (85.78 - 90.78)		21.55 (18.32 - 25.17)		17.68 (15.21 - 20.45)	
50.00	8.13 (6.43 - 10.24)		81.16 (78.66 - 83.43)		7.87 (6.30 - 9.79)		5.94 (4.19 - 8.35)	
Spam Misc%	Ham Misc%							
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.05	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.10	98.41 (94.96 - 99.51)		99.00 (98.72 - 99.23)		97.26 (94.98 - 98.52)		99.65 (99.06 - 99.87)	
0.20	95.70 (88.49 - 98.47)		98.96 (98.65 - 99.19)		95.84 (92.04 - 97.87)		99.63 (98.45 - 99.91)	
0.50	94.53 (89.36 - 97.26)		98.92 (98.60 - 99.17)		94.35 (90.26 - 96.78)		98.80 (94.44 - 99.75)	
1.00	91.56 (84.48 - 95.58)		98.84 (98.38 - 99.18)		88.77 (81.07 - 93.58)		96.32 (90.01 - 98.71)	
2.00	84.18 (79.93 - 87.66)		98.28 (97.45 - 98.85)		86.28 (82.85 - 89.11)		87.61 (74.78 - 94.40)	
5.00	72.09 (64.71 - 78.44)		95.86 (93.51 - 97.38)		74.80 (61.78 - 84.50)		58.53 (43.04 - 72.50)	
10.00	45.50 (39.86 - 51.25)		92.68 (91.89 - 93.40)		43.57 (37.24 - 50.12)		35.79 (29.97 - 42.06)	
20.00	25.49 (22.47 - 28.76)		87.40 (85.48 - 89.10)		21.23 (17.27 - 25.82)		15.01 (8.26 - 25.72)	
50.00	1.59 (1.28 - 1.98)		76.81 (74.98 - 78.54)		1.57 (1.32 - 1.87)		0.64 (0.47 - 0.88)	

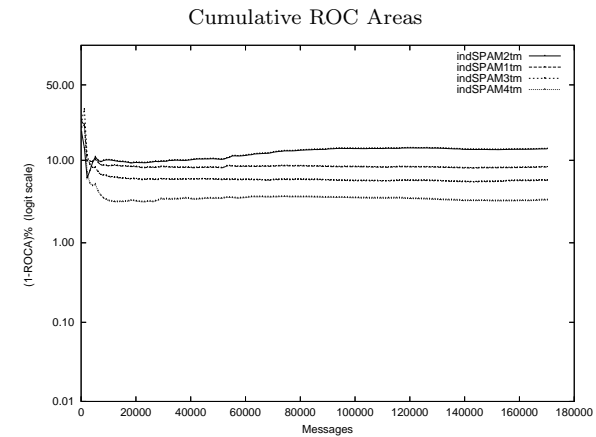
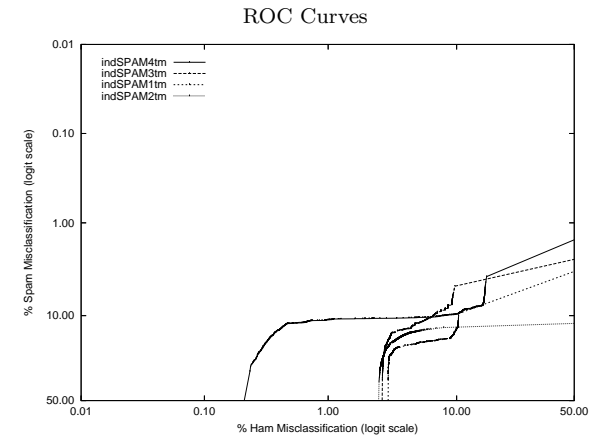


Learning Curves

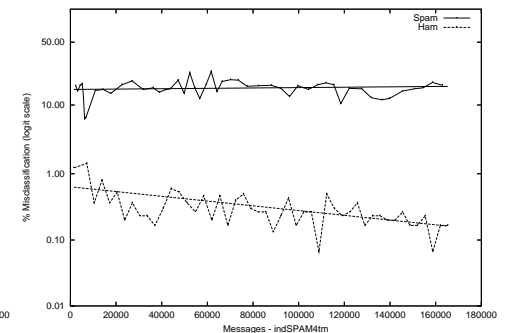
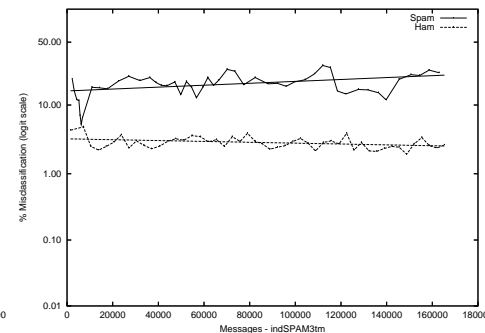
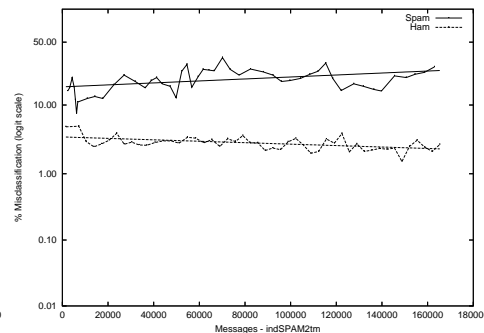
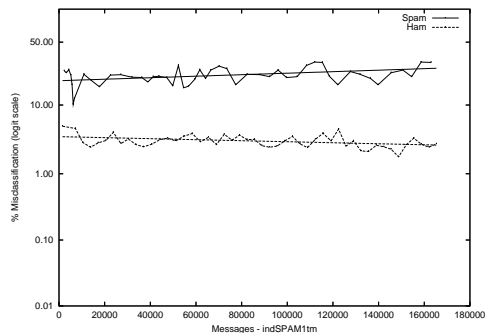


ind - Indiana U. (Yang) on Private T.M. Corpus

Gold Std.	indSPAM1tm		indSPAM2tm		indSPAM3tm		indSPAM4tm					
	ham	spam	ham	spam	ham	spam	ham	spam				
Filter Classification	ham spam	146047 4638	4693 14823	ham spam	146356 4329	4178 15338	ham spam	146285 4400	3704 15812	ham spam	150177 508	3267 16249
Ham Misc%	3.08 (2.99-3.17)		2.87 (2.79-2.96)		2.92 (2.84-3.01)		0.34 (0.31-0.37)					
Spam Misc%	24.05 (23.45-24.65)		21.41 (20.83-21.99)		18.98 (18.43-19.54)		16.74 (16.22-17.27)					
Logit av Misc%	9.11 (8.93-9.30)		8.24 (8.07-8.41)		7.74 (7.58-7.91)		2.54 (2.43-2.66)					
(1-ROCA)%	8.382 (8.140 - 8.630)		13.462 (13.060 - 13.875)		5.843 (5.622 - 6.072)		3.388 (3.209 - 3.577)					
Ham Misc%	Spam Misc%											
0.01	99.82 (99.72 - 99.88)		99.85 (99.76 - 99.91)		99.83 (99.75 - 99.89)		99.68 (99.43 - 99.82)					
0.02	99.77 (99.66 - 99.84)		99.82 (99.71 - 99.88)		99.81 (99.72 - 99.86)		99.42 (99.10 - 99.62)					
0.05	99.62 (99.50 - 99.71)		99.71 (99.59 - 99.79)		99.67 (99.57 - 99.75)		98.33 (97.90 - 98.67)					
0.10	99.44 (99.30 - 99.55)		99.44 (99.28 - 99.56)		99.41 (99.26 - 99.53)		96.77 (96.17 - 97.27)					
0.20	98.88 (98.69 - 99.05)		98.92 (98.72 - 99.09)		98.87 (98.66 - 99.04)		57.06 (37.04 - 75.00)					
0.50	97.41 (97.12 - 97.66)		97.44 (97.16 - 97.69)		97.28 (97.01 - 97.53)		11.99 (11.51 - 12.49)					
1.00	94.59 (94.22 - 94.94)		94.57 (94.13 - 94.99)		94.41 (94.04 - 94.77)		11.04 (10.57 - 11.53)					
2.00	89.02 (88.50 - 89.52)		89.60 (89.01 - 90.16)		89.00 (88.42 - 89.56)		10.71 (10.28 - 11.16)					
5.00	18.76 (18.11 - 19.43)		14.42 (13.96 - 14.90)		12.01 (11.52 - 12.51)		10.49 (10.05 - 10.95)					
10.00	13.74 (13.05 - 14.46)		13.10 (12.65 - 13.56)		4.91 (4.60 - 5.25)		9.66 (9.25 - 10.09)					
20.00	8.48 (8.12 - 8.85)		12.85 (12.40 - 13.30)		4.39 (4.09 - 4.70)		3.71 (3.46 - 3.99)					
50.00	5.40 (5.11 - 5.69)		12.07 (11.64 - 12.52)		3.08 (2.84 - 3.35)		2.39 (2.18 - 2.62)					
Spam Misc%	Ham Misc%											
0.01	99.46 (89.91 - 99.97)		99.66 (99.59 - 99.72)		99.83 (96.97 - 99.99)		99.84 (94.18 - 100.00)					
0.02	99.34 (97.64 - 99.82)		99.61 (99.55 - 99.67)		99.45 (98.49 - 99.80)		99.71 (98.37 - 99.95)					
0.05	99.19 (99.12 - 99.25)		99.59 (99.54 - 99.64)		99.22 (99.07 - 99.35)		98.45 (97.03 - 99.19)					
0.10	99.15 (99.08 - 99.22)		99.49 (99.33 - 99.61)		99.15 (99.06 - 99.24)		97.30 (95.85 - 98.25)					
0.20	98.80 (98.38 - 99.12)		99.20 (99.09 - 99.30)		98.84 (98.58 - 99.05)		94.59 (92.87 - 95.92)					
0.50	97.91 (97.73 - 98.07)		98.60 (98.03 - 99.00)		98.05 (97.79 - 98.29)		88.57 (86.43 - 90.41)					
1.00	96.56 (93.91 - 98.08)		97.44 (97.21 - 97.66)		96.47 (95.94 - 96.93)		78.52 (75.54 - 81.23)					
2.00	82.12 (80.44 - 83.69)		96.89 (96.69 - 97.08)		73.65 (68.91 - 77.91)		57.57 (53.02 - 61.99)					
5.00	53.95 (51.35 - 56.54)		92.52 (91.85 - 93.14)		9.70 (7.32 - 12.75)		15.94 (15.75 - 16.14)					
10.00	10.33 (10.04 - 10.64)		89.28 (89.09 - 89.48)		6.76 (6.46 - 7.08)		8.12 (5.88 - 11.11)					
20.00	3.82 (3.48 - 4.20)		3.01 (2.90 - 3.13)		2.83 (2.75 - 2.92)		0.30 (0.27 - 0.33)					
50.00	3.00 (2.92 - 3.08)		2.53 (2.45 - 2.61)		2.69 (2.61 - 2.76)		0.21 (0.19 - 0.24)					



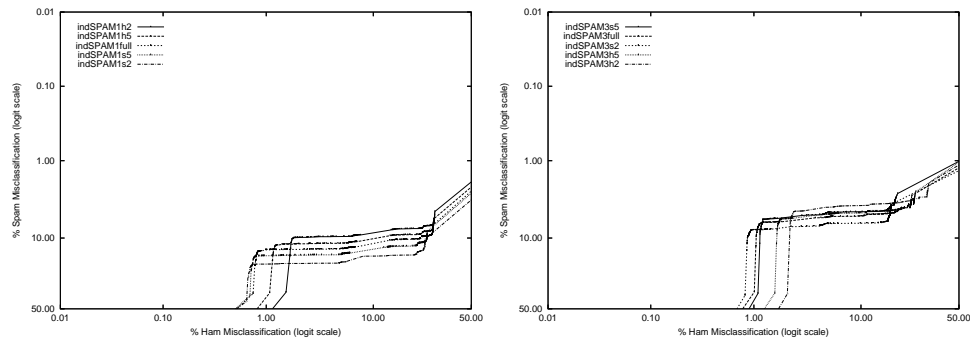
Learning Curves



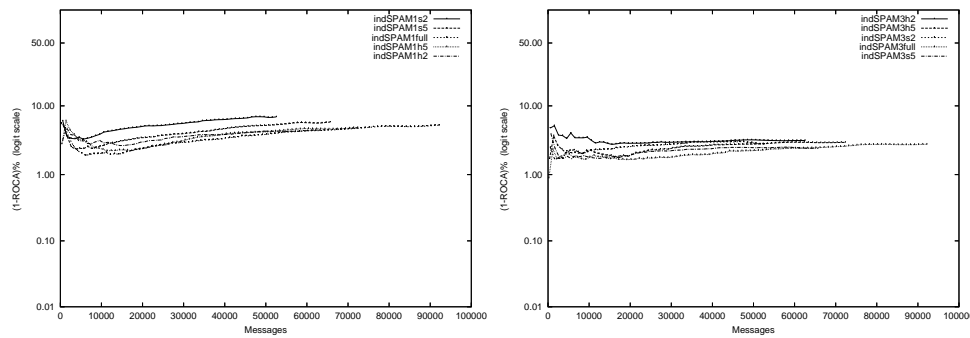
ind - Indiana U. (Yang) - Public subset Corpora

	indSPAM1full	indSPAM1h2	indSPAM1h5	indSPAM1s2	indSPAM1s5
Ham Misc%	0.82 (0.73-0.91)	1.75 (1.50-2.03)	1.20 (1.05-1.36)	0.70 (0.62-0.79)	0.75 (0.67-0.84)
Spam Misc%	15.16 (14.86-15.47)	11.02 (10.75-11.29)	13.11 (12.82-13.40)	21.48 (20.78-22.19)	17.58 (17.12-18.04)
Logit av Misc%	3.70 (3.50-3.90)	4.49 (4.17-4.83)	4.10 (3.85-4.37)	4.21 (3.96-4.47)	3.86 (3.65-4.09)
(1-ROCA)%	5.346 (5.205 - 5.491)	4.576 (4.392 - 4.767)	4.939 (4.780 - 5.103)	7.053 (6.770 - 7.347)	5.951 (5.738 - 6.173)
	indSPAM3full	indSPAM3h2	indSPAM3h5	indSPAM3s2	indSPAM3s5
Ham Misc%	1.09 (0.99-1.19)	2.27 (1.98-2.58)	1.70 (1.52-1.89)	0.89 (0.80-0.99)	1.18 (1.08-1.29)
Spam Misc%	7.66 (7.43-7.89)	5.56 (5.36-5.75)	6.95 (6.73-7.17)	9.32 (8.83-9.83)	7.02 (6.71-7.33)
Logit av Misc%	2.93 (2.79-3.08)	3.56 (3.33-3.81)	3.46 (3.28-3.66)	2.95 (2.78-3.13)	2.92 (2.78-3.07)
(1-ROCA)%	2.822 (2.710 - 2.940)	3.210 (2.977 - 3.461)	3.012 (2.863 - 3.168)	2.844 (2.672 - 3.027)	2.471 (2.345 - 2.604)

ROC Curves

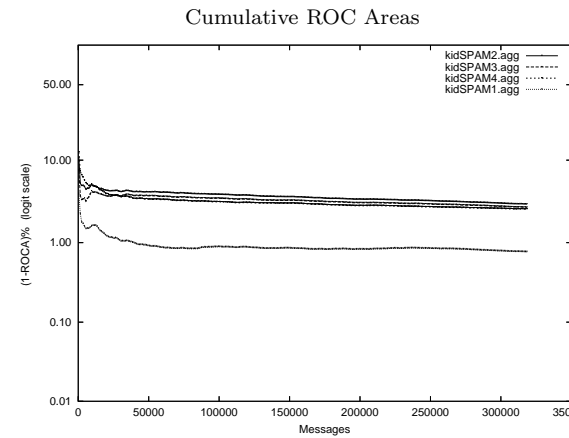
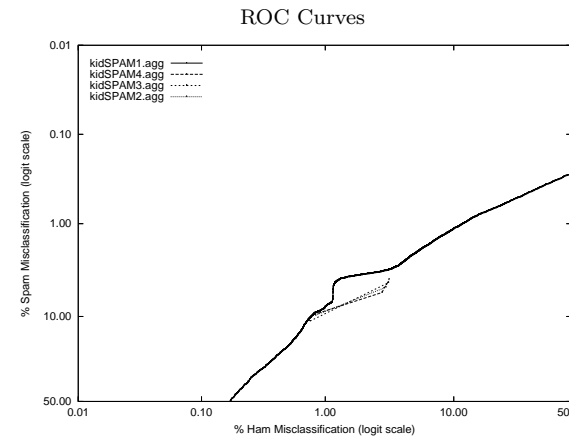


Cumulative ROC Area

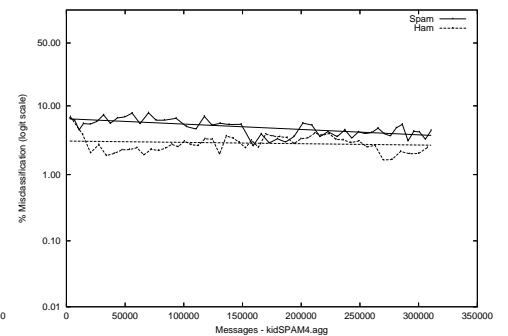
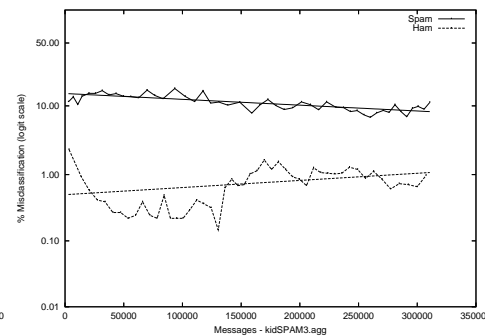
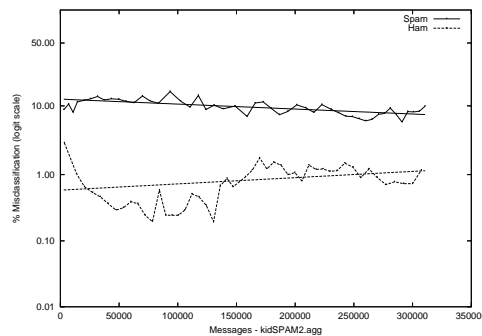
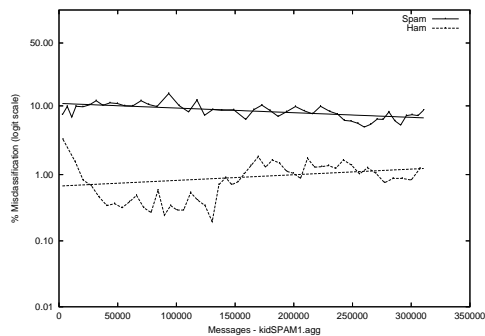


kid - Beijing U. of Posts and Telecom. on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	kidSPAM1.agg		kidSPAM2.agg		kidSPAM3.agg		kidSPAM4.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	203445	9733	203637	10980	203805	12568	199310	5714
	1907	103395	1715	102148	1547	100559	6042	107414
Ham Misc%	0.93 (0.89-0.97)		0.84 (0.80-0.88)		0.75 (0.72-0.79)		2.94 (2.87-3.02)	
Spam Misc%	8.60 (8.44-8.77)		9.71 (9.53-9.88)		11.11 (10.93-11.29)		5.05 (4.92-5.18)	
Logit av Misc%	2.88 (2.82-2.96)		2.92 (2.85-2.99)		2.99 (2.91-3.07)		3.86 (3.79-3.93)	
(1-ROCA)%	0.768 (0.737 - 0.799)		3.003 (2.921 - 3.088)		2.741 (2.672 - 2.812)		2.606 (2.537 - 2.676)	
Ham Misc%	Spam Misc%							
0.01	96.34 (91.97 - 98.38)		99.01 (98.41 - 99.38)		98.84 (98.21 - 99.25)		99.01 (98.44 - 99.37)	
0.02	89.97 (87.99 - 91.65)		97.68 (96.83 - 98.31)		97.78 (97.00 - 98.37)		97.68 (96.94 - 98.24)	
0.05	79.32 (76.11 - 82.20)		94.65 (93.51 - 95.60)		94.51 (93.24 - 95.56)		95.01 (93.91 - 95.92)	
0.10	66.13 (62.20 - 69.86)		88.29 (86.64 - 89.77)		88.23 (86.48 - 89.78)		89.15 (87.54 - 90.57)	
0.20	43.59 (40.16 - 47.09)		77.60 (75.22 - 79.81)		75.86 (73.62 - 77.97)		78.31 (76.07 - 80.40)	
0.50	19.46 (18.16 - 20.84)		47.03 (43.60 - 50.49)		43.18 (39.52 - 46.92)		48.74 (45.63 - 51.86)	
1.00	7.96 (7.46 - 8.49)		9.36 (9.17 - 9.55)		10.44 (10.22 - 10.67)		9.03 (8.85 - 9.21)	
2.00	3.62 (3.51 - 3.73)		7.19 (7.00 - 7.39)		7.69 (7.47 - 7.91)		7.20 (7.02 - 7.39)	
5.00	2.18 (2.09 - 2.26)		4.84 (4.72 - 4.97)		4.34 (4.22 - 4.47)		3.99 (3.87 - 4.11)	
10.00	1.14 (1.08 - 1.21)		4.58 (4.46 - 4.71)		4.09 (3.97 - 4.21)		3.78 (3.67 - 3.90)	
20.00	0.66 (0.61 - 0.70)		4.08 (3.95 - 4.20)		3.64 (3.53 - 3.75)		3.36 (3.25 - 3.46)	
50.00	0.28 (0.25 - 0.31)		2.51 (2.42 - 2.61)		2.28 (2.19 - 2.37)		2.08 (2.01 - 2.17)	
Spam Misc%	Ham Misc%							
0.01	96.79 (95.28 - 97.83)		99.84 (99.70 - 99.91)		99.68 (99.40 - 99.83)		99.83 (99.69 - 99.91)	
0.02	95.18 (93.26 - 96.57)		99.60 (99.37 - 99.74)		99.37 (99.04 - 99.59)		99.55 (99.34 - 99.69)	
0.05	90.70 (87.92 - 92.89)		99.14 (98.86 - 99.35)		98.77 (98.40 - 99.05)		98.94 (98.63 - 99.17)	
0.10	79.98 (75.58 - 83.76)		98.17 (97.80 - 98.48)		97.83 (97.40 - 98.19)		97.81 (97.41 - 98.15)	
0.20	62.98 (59.23 - 66.57)		96.14 (95.55 - 96.65)		95.62 (95.06 - 96.13)		95.57 (94.99 - 96.08)	
0.50	28.16 (25.57 - 30.90)		89.95 (89.04 - 90.78)		88.91 (87.97 - 89.78)		87.89 (86.92 - 88.80)	
1.00	11.66 (11.02 - 12.32)		80.28 (78.97 - 81.53)		77.81 (76.58 - 78.98)		76.26 (74.89 - 77.58)	
2.00	5.48 (5.24 - 5.72)		59.99 (58.28 - 61.69)		56.05 (54.25 - 57.84)		51.91 (50.09 - 53.72)	
5.00	1.16 (1.11 - 1.21)		3.02 (2.39 - 3.82)		2.98 (2.90 - 3.07)		2.96 (2.88 - 3.04)	
10.00	0.77 (0.73 - 0.82)		0.83 (0.80 - 0.87)		1.17 (1.08 - 1.25)		0.86 (0.82 - 0.90)	
20.00	0.49 (0.46 - 0.52)		0.73 (0.70 - 0.77)		0.68 (0.64 - 0.71)		0.76 (0.73 - 0.80)	
50.00	0.17 (0.15 - 0.19)		0.47 (0.44 - 0.50)		0.44 (0.41 - 0.47)		0.49 (0.46 - 0.52)	

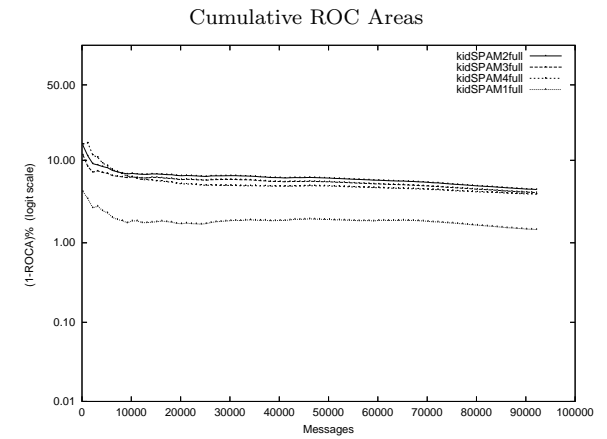
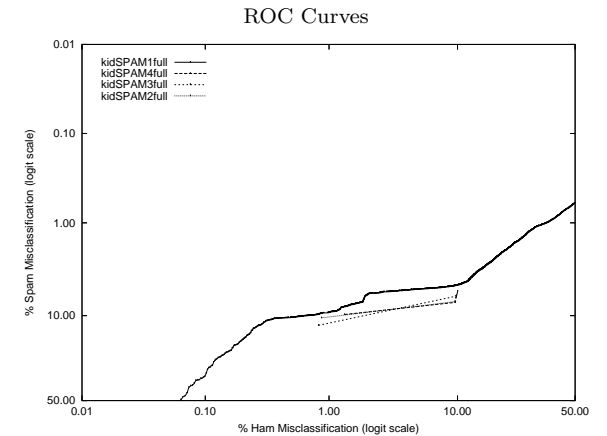


Learning Curves

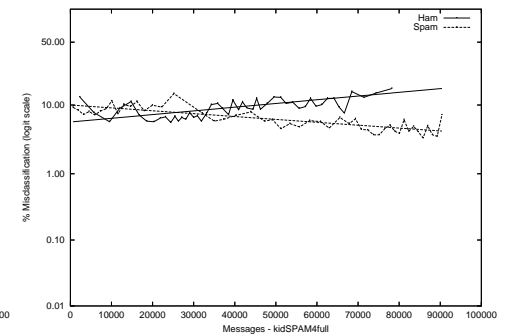
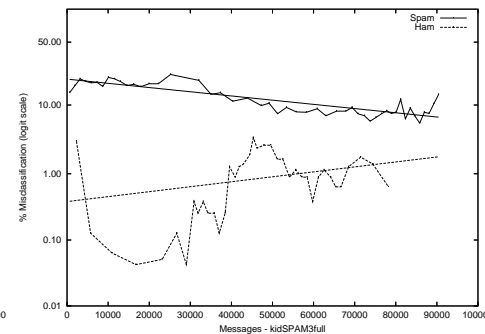
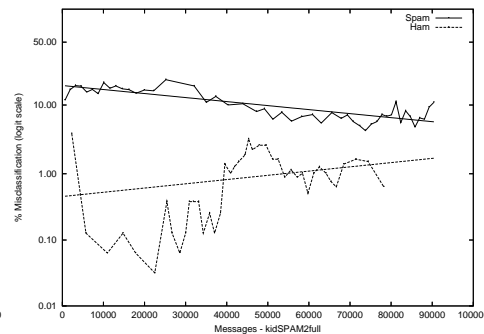
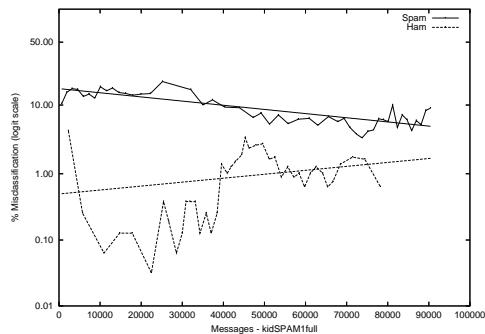


kid - Beijing U. of Posts and Telecom. on Public trec05p-1/full Corpus

Gold Std. Filter Classification	kidSPAM1full		kidSPAM2full		kidSPAM3full		kidSPAM4full	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	39042 357	4960 47830	39056 343	5557 47233	39075 324	6591 46198	35563 3836	3470 49320
Ham Misc%	0.91 (0.81-1.00)		0.87 (0.78-0.97)		0.82 (0.74-0.92)		9.74 (9.45-10.03)	
Spam Misc%	9.40 (9.15-9.65)		10.53 (10.27-10.79)		12.49 (12.20-12.77)		6.57 (6.36-6.79)	
Logit av Misc%	2.99 (2.83-3.15)		3.11 (2.95-3.28)		3.33 (3.15-3.51)		8.01 (7.84-8.19)	
(1-ROCA)%	1.463 (1.399 - 1.529)		4.544 (4.408 - 4.685)		4.167 (4.031 - 4.308)		3.990 (3.840 - 4.145)	
Ham Misc%	Spam Misc%							
0.01	97.89 (92.76 - 99.41)		98.48 (96.11 - 99.41)		98.53 (95.49 - 99.53)		98.80 (96.80 - 99.55)	
0.02	88.96 (69.17 - 96.66)		98.15 (96.31 - 99.09)		98.26 (96.52 - 99.14)		98.43 (96.94 - 99.20)	
0.05	58.12 (35.01 - 78.14)		96.00 (93.47 - 97.58)		95.52 (92.87 - 97.22)		96.94 (95.30 - 98.02)	
0.10	34.93 (25.48 - 45.74)		91.65 (88.24 - 94.14)		90.62 (87.35 - 93.12)		93.74 (91.56 - 95.39)	
0.20	17.36 (14.24 - 21.01)		81.13 (75.90 - 85.44)		80.65 (75.86 - 84.68)		86.81 (83.90 - 89.26)	
0.50	10.46 (10.16 - 10.76)		51.47 (43.41 - 59.45)		45.11 (38.37 - 52.04)		66.11 (61.37 - 70.55)	
1.00	9.34 (9.06 - 9.63)		10.47 (10.20 - 10.74)		12.36 (12.09 - 12.63)		32.76 (26.25 - 40.01)	
2.00	6.16 (5.61 - 6.76)		10.13 (9.86 - 10.39)		11.66 (11.38 - 11.94)		9.57 (9.31 - 9.84)	
5.00	5.34 (5.15 - 5.52)		8.90 (8.65 - 9.16)		9.57 (9.28 - 9.86)		8.65 (8.40 - 8.90)	
10.00	4.81 (4.63 - 5.00)		7.14 (6.93 - 7.36)		6.28 (6.07 - 6.49)		5.75 (5.06 - 6.51)	
20.00	2.20 (2.05 - 2.35)		6.37 (6.17 - 6.58)		5.62 (5.42 - 5.82)		4.98 (4.77 - 5.20)	
50.00	0.61 (0.54 - 0.68)		4.04 (3.88 - 4.21)		3.59 (3.43 - 3.75)		3.17 (3.00 - 3.35)	
Spam Misc%	Ham Misc%							
0.01	95.64 (93.54 - 97.08)		99.77 (99.47 - 99.90)		99.69 (99.33 - 99.86)		99.82 (99.56 - 99.92)	
0.02	93.53 (90.69 - 95.54)		99.64 (99.31 - 99.81)		99.56 (99.22 - 99.76)		99.66 (99.37 - 99.82)	
0.05	89.13 (85.54 - 91.92)		99.29 (98.96 - 99.52)		99.07 (98.67 - 99.34)		98.97 (98.60 - 99.25)	
0.10	82.25 (78.98 - 85.10)		98.53 (98.09 - 98.87)		98.07 (97.55 - 98.48)		97.85 (97.20 - 98.34)	
0.20	72.71 (69.10 - 76.04)		97.21 (96.66 - 97.67)		96.85 (96.23 - 97.37)		96.11 (95.25 - 96.82)	
0.50	53.92 (50.97 - 56.84)		93.12 (92.17 - 93.97)		92.42 (91.45 - 93.28)		90.66 (89.51 - 91.69)	
1.00	37.16 (34.08 - 40.36)		87.31 (86.00 - 88.51)		85.77 (84.51 - 86.94)		83.43 (81.96 - 84.81)	
2.00	21.58 (20.41 - 22.79)		75.49 (73.89 - 77.02)		72.42 (70.71 - 74.07)		68.33 (66.31 - 70.29)	
5.00	8.63 (7.19 - 10.34)		37.78 (35.52 - 40.09)		28.19 (25.61 - 30.92)		19.70 (16.38 - 23.51)	
10.00	0.69 (0.55 - 0.85)		2.40 (1.78 - 3.22)		4.31 (3.91 - 4.76)		1.33 (1.17 - 1.51)	
20.00	0.17 (0.13 - 0.22)		0.78 (0.69 - 0.89)		0.76 (0.69 - 0.84)		1.17 (1.07 - 1.28)	
50.00	0.06 (0.04 - 0.10)		0.52 (0.45 - 0.61)		0.45 (0.40 - 0.52)		0.77 (0.70 - 0.86)	

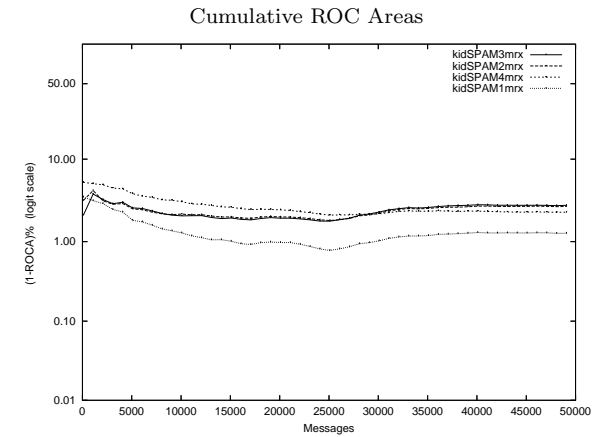
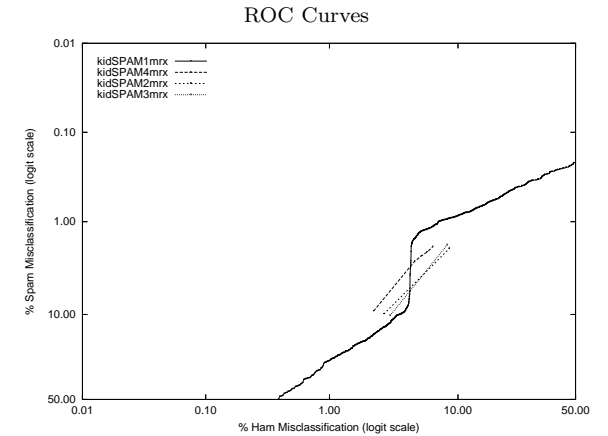


Learning Curves

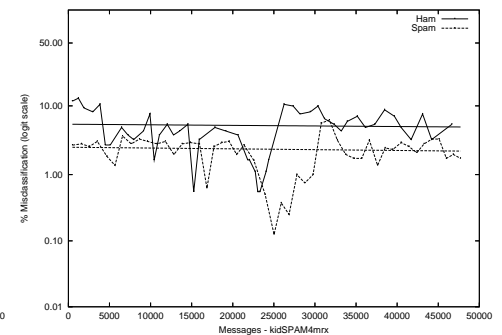
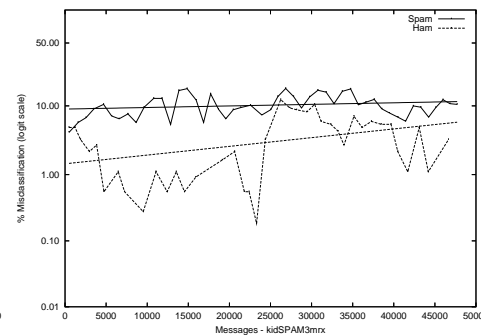
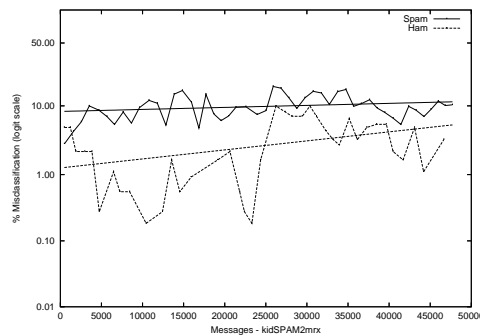
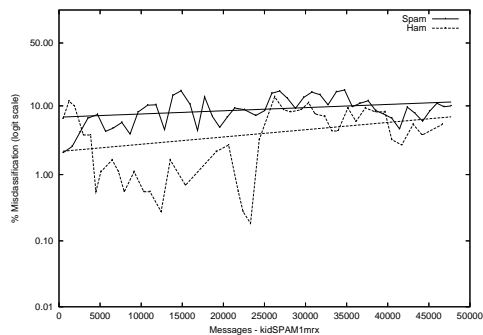


kid - Beijing U. of Posts and Telecom. on Private Mr X. Corpus

Gold Std. Filter Classification	kidSPAM1mrx		kidSPAM2mrx		kidSPAM3mrx		kidSPAM4mrx	
	ham	spam	ham	spam	ham	spam	ham	spam
Filter	8675	3646	8793	3961	8764	4113	8558	957
Classification	363	36402	245	36087	274	35935	480	39091
Ham Misc%	4.02 (3.62-4.44)		2.71 (2.39-3.07)		3.03 (2.69-3.41)		5.31 (4.86-5.79)	
Spam Misc%	9.10 (8.82-9.39)		9.89 (9.60-10.19)		10.27 (9.97-10.57)		2.39 (2.24-2.54)	
Logit av Misc%	6.08 (5.77-6.40)		5.24 (4.92-5.58)		5.64 (5.32-5.99)		3.57 (3.39-3.77)	
(1-ROCA)%	1.274 (1.150 - 1.412)		2.738 (2.506 - 2.992)		2.822 (2.612 - 3.049)		2.326 (2.122 - 2.548)	
Ham Misc%	Spam Misc%							
0.01	99.38 (97.36 - 99.86)		99.63 (nan - nan)		99.64 (96.24 - 99.97)		99.84 (nan - nan)	
0.02	98.17 (89.53 - 99.70)		99.21 (94.82 - 99.88)		99.22 (96.02 - 99.85)		99.62 (98.07 - 99.93)	
0.05	89.30 (59.57 - 97.93)		98.93 (97.06 - 99.61)		98.50 (96.01 - 99.45)		98.92 (97.12 - 99.60)	
0.10	83.55 (69.22 - 91.98)		97.64 (95.47 - 98.78)		97.67 (95.90 - 98.68)		98.23 (96.24 - 99.17)	
0.20	67.46 (51.95 - 79.90)		93.67 (90.06 - 96.03)		93.87 (90.65 - 96.03)		94.89 (91.87 - 96.82)	
0.50	43.34 (36.97 - 49.94)		81.91 (75.50 - 86.93)		83.69 (78.47 - 87.84)		82.27 (75.66 - 87.38)	
1.00	26.95 (23.25 - 30.99)		66.30 (59.12 - 72.81)		71.26 (65.11 - 76.72)		61.39 (53.06 - 69.10)	
2.00	17.62 (15.88 - 19.51)		34.21 (24.73 - 45.14)		41.67 (33.29 - 50.56)		20.33 (10.41 - 35.91)	
5.00	1.39 (1.18 - 1.64)		6.99 (6.23 - 7.84)		6.97 (6.21 - 7.81)		2.60 (2.22 - 3.03)	
10.00	0.87 (0.78 - 0.97)		1.95 (1.82 - 2.09)		1.78 (1.64 - 1.92)		1.82 (1.71 - 1.93)	
20.00	0.52 (0.45 - 0.61)		1.70 (1.57 - 1.84)		1.53 (1.41 - 1.65)		1.57 (1.46 - 1.68)	
50.00	0.22 (0.18 - 0.27)		1.08 (0.98 - 1.19)		0.99 (0.90 - 1.08)		0.98 (0.88 - 1.08)	
Spam Misc%	Ham Misc%							
0.01	89.60 (83.18 - 93.75)		99.49 (98.62 - 99.81)		99.23 (97.87 - 99.72)		99.02 (97.61 - 99.60)	
0.02	86.87 (82.18 - 90.46)		98.96 (97.88 - 99.49)		98.95 (97.79 - 99.50)		98.36 (96.86 - 99.15)	
0.05	80.83 (76.83 - 84.27)		97.46 (95.90 - 98.43)		97.49 (96.11 - 98.39)		96.94 (95.38 - 97.98)	
0.10	72.76 (67.88 - 77.15)		95.49 (93.98 - 96.63)		94.73 (92.86 - 96.14)		94.49 (92.60 - 95.92)	
0.20	52.95 (44.66 - 61.09)		91.41 (89.32 - 93.13)		90.17 (87.73 - 92.18)		89.67 (86.89 - 91.91)	
0.50	20.81 (18.01 - 23.92)		79.21 (76.12 - 82.00)		75.35 (71.87 - 78.52)		76.45 (72.21 - 80.23)	
1.00	7.23 (6.17 - 8.44)		55.73 (50.95 - 60.41)		49.57 (44.84 - 54.31)		48.81 (43.80 - 53.83)	
2.00	4.44 (4.04 - 4.87)		8.61 (6.43 - 11.44)		8.12 (7.55 - 8.73)		6.28 (5.68 - 6.94)	
5.00	4.35 (3.95 - 4.79)		6.43 (5.86 - 7.05)		6.36 (5.84 - 6.93)		3.85 (3.46 - 4.29)	
10.00	3.64 (3.24 - 4.08)		2.71 (2.37 - 3.10)		3.19 (2.84 - 3.58)		2.25 (1.95 - 2.59)	
20.00	1.59 (1.36 - 1.87)		2.42 (2.10 - 2.79)		2.73 (2.41 - 3.09)		2.00 (1.73 - 2.32)	
50.00	0.39 (0.28 - 0.53)		1.46 (1.23 - 1.74)		1.68 (1.45 - 1.95)		1.28 (1.07 - 1.54)	

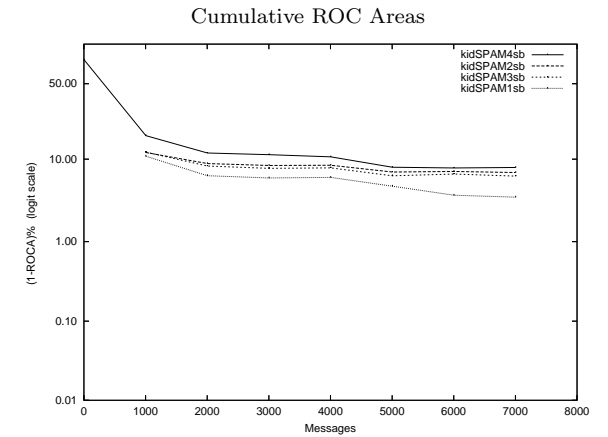
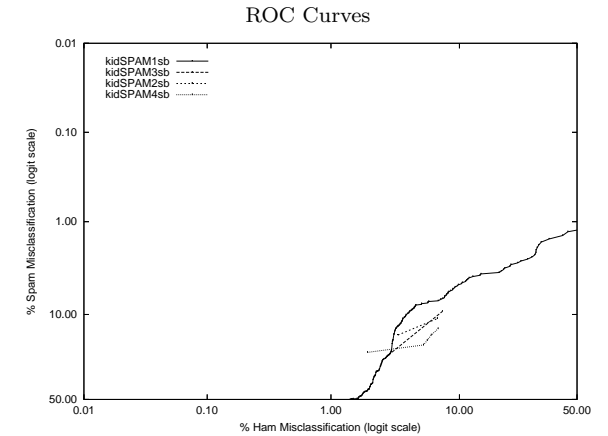


Learning Curves

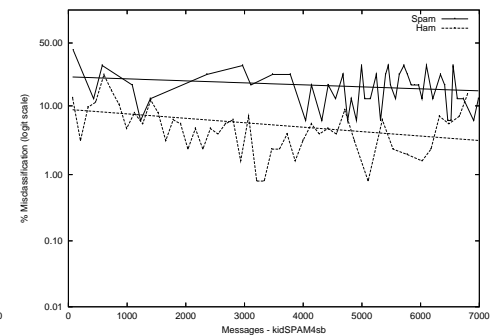
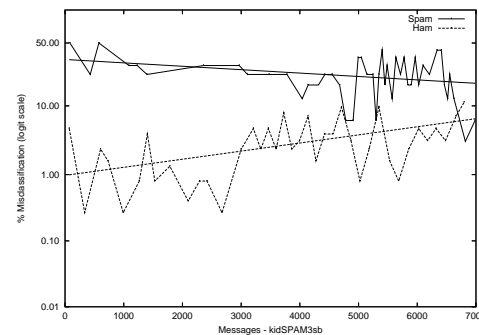
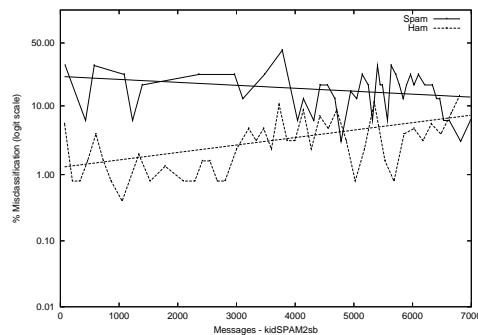
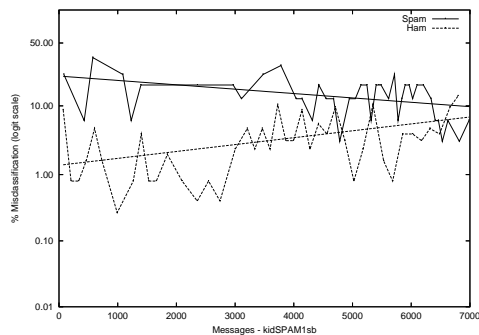


kid - Beijing U. of Posts and Telecom. on Private S.B. Corpus

Gold Std. Filter Classification	kidSPAM1sb			kidSPAM2sb			kidSPAM3sb			kidSPAM4sb		
	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	
Ham Misc%	3.37 (2.94-3.85)			3.40 (2.97-3.88)			2.86 (2.46-3.30)			5.75 (5.18-6.35)		
Spam Misc%	13.57 (11.23-16.18)			16.15 (13.63-18.93)			24.42 (21.43-27.60)			18.09 (15.44-20.98)		
Logit av Misc%	6.89 (6.14-7.73)			7.62 (6.83-8.49)			8.89 (8.03-9.83)			10.40 (9.46-11.43)		
(1-ROCA)%	3.553 (2.826 - 4.460)			7.020 (5.919 - 8.308)			6.360 (5.247 - 7.690)			8.042 (6.843 - 9.430)		
Ham Misc%	Spam Misc%											
0.01	100.00 (nan - nan)			98.84 (nan - nan)			98.97 (nan - nan)			98.84 (nan - nan)		
0.02	100.00 (nan - nan)			98.84 (nan - nan)			98.71 (nan - nan)			98.84 (95.47 - 99.71)		
0.05	100.00 (nan - nan)			98.58 (95.93 - 99.51)			97.55 (93.37 - 99.12)			98.19 (94.88 - 99.38)		
0.10	99.22 (nan - nan)			97.29 (94.14 - 98.77)			93.67 (87.48 - 96.91)			95.22 (90.12 - 97.75)		
0.20	95.99 (84.40 - 99.07)			93.15 (88.44 - 96.03)			91.99 (87.33 - 95.03)			92.64 (87.06 - 95.92)		
0.50	77.78 (62.42 - 88.06)			84.63 (78.08 - 89.48)			81.01 (73.97 - 86.49)			78.42 (70.51 - 84.67)		
1.00	57.36 (52.24 - 62.34)			73.39 (66.42 - 79.35)			69.51 (61.82 - 76.24)			59.95 (50.44 - 68.76)		
2.00	43.67 (34.56 - 53.22)			47.67 (38.90 - 56.59)			44.19 (34.79 - 54.01)			22.87 (17.52 - 29.26)		
5.00	8.01 (6.10 - 10.45)			13.57 (11.33 - 16.16)			18.09 (14.99 - 21.67)			20.03 (17.61 - 22.68)		
10.00	4.91 (3.54 - 6.77)			10.47 (8.42 - 12.93)			8.91 (7.11 - 11.12)			13.31 (11.09 - 15.88)		
20.00	3.49 (2.27 - 5.32)			8.91 (7.07 - 11.19)			7.49 (5.85 - 9.55)			10.85 (8.88 - 13.19)		
50.00	1.29 (0.64 - 2.58)			5.30 (4.01 - 6.97)			4.65 (3.20 - 6.71)			6.98 (5.29 - 9.14)		
Spam Misc%	Ham Misc%											
0.01	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.10	96.53 (89.35 - 98.93)			93.87 (76.95 - 98.60)			93.95 (78.46 - 98.51)			97.56 (79.33 - 99.76)		
0.20	94.03 (79.07 - 98.50)			92.83 (77.42 - 97.99)			92.89 (80.28 - 97.67)			96.61 (86.63 - 99.21)		
0.50	83.95 (64.52 - 93.77)			91.89 (82.16 - 96.54)			91.96 (80.96 - 96.85)			92.71 (84.53 - 96.73)		
1.00	79.74 (45.80 - 94.83)			84.85 (75.43 - 91.08)			85.96 (72.94 - 93.29)			88.15 (79.32 - 93.52)		
2.00	32.22 (19.07 - 48.94)			76.84 (66.42 - 84.76)			77.11 (63.64 - 86.64)			82.26 (73.38 - 88.64)		
5.00	9.84 (6.11 - 15.48)			55.54 (44.02 - 66.49)			47.08 (31.84 - 62.88)			63.92 (51.62 - 74.63)		
10.00	4.04 (3.22 - 5.06)			13.15 (4.57 - 32.34)			7.03 (3.78 - 12.72)			26.77 (16.16 - 40.95)		
20.00	3.07 (2.66 - 3.53)			3.29 (2.87 - 3.77)			4.48 (3.52 - 5.68)			5.10 (3.11 - 8.26)		
50.00	1.43 (1.04 - 1.96)			1.89 (1.58 - 2.27)			1.73 (1.39 - 2.16)			1.25 (1.01 - 1.55)		

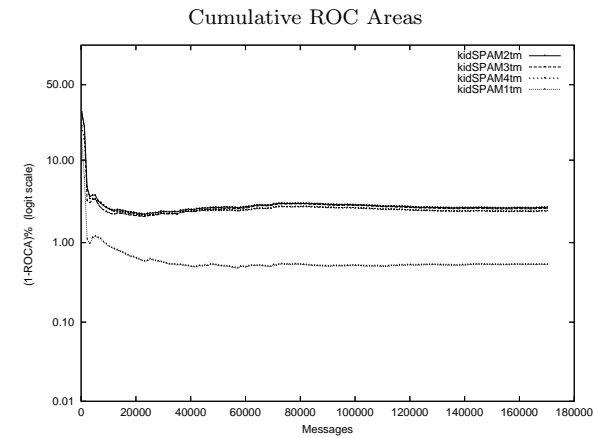
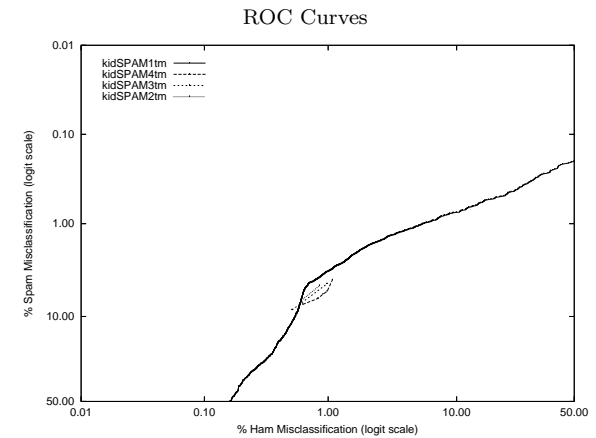


Learning Curves

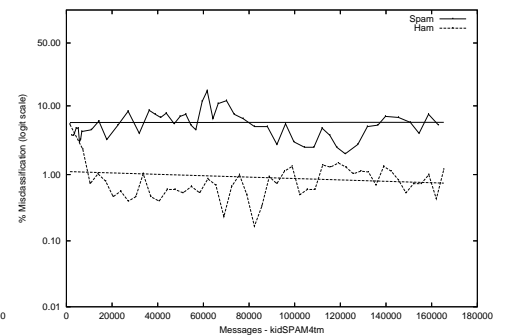
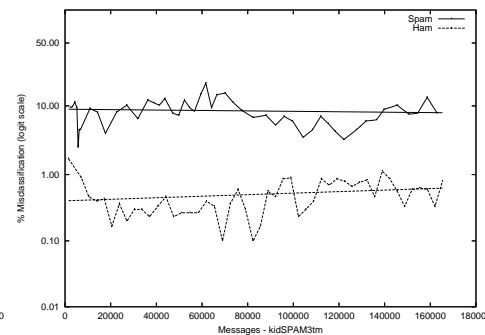
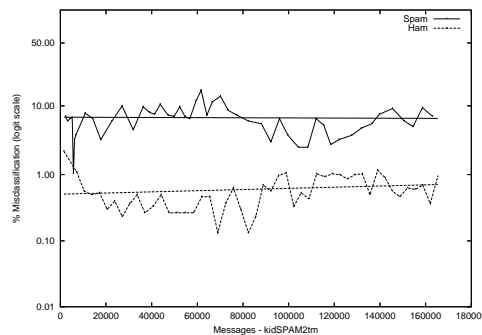
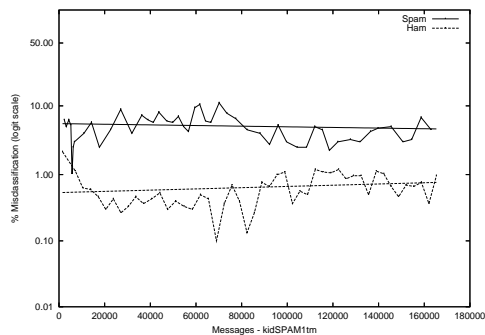


kid - Beijing U. of Posts and Telecom. on Private T.M. Corpus

Gold Std. Filter Classification	kidSPAM1tm		kidSPAM2tm		kidSPAM3tm		kidSPAM4tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	149708	1022	149770	1337	149914	1675	149317	1147
	977	18494	915	18179	771	17841	1368	18369
Ham Misc%	0.65 (0.61-0.69)		0.61 (0.57-0.65)		0.51 (0.48-0.55)		0.91 (0.86-0.96)	
Spam Misc%	5.24 (4.93-5.56)		6.85 (6.50-7.21)		8.58 (8.19-8.98)		5.88 (5.55-6.22)	
Logit av Misc%	1.86 (1.78-1.95)		2.08 (1.99-2.16)		2.15 (2.06-2.24)		2.34 (2.25-2.43)	
(1-ROCA)%	0.530 (0.477 - 0.589)		2.749 (2.601 - 2.906)		2.653 (2.482 - 2.836)		2.473 (2.320 - 2.636)	
Ham Misc%	Spam Misc%							
0.01	91.28 (88.74 - 93.30)		98.63 (97.45 - 99.27)		98.41 (97.14 - 99.12)		98.44 (97.38 - 99.07)	
0.02	87.01 (81.65 - 90.97)		97.26 (95.83 - 98.21)		96.94 (95.53 - 97.91)		96.98 (95.70 - 97.88)	
0.05	73.96 (71.16 - 76.59)		93.00 (91.04 - 94.55)		92.20 (90.00 - 93.94)		92.99 (91.16 - 94.47)	
0.10	62.56 (58.97 - 66.02)		85.16 (82.29 - 87.64)		82.11 (79.06 - 84.80)		85.34 (82.93 - 87.46)	
0.20	38.73 (33.35 - 44.41)		70.20 (66.27 - 73.85)		65.01 (61.21 - 68.64)		70.44 (67.19 - 73.50)	
0.50	12.05 (10.31 - 14.02)		23.71 (18.60 - 29.72)		10.39 (6.42 - 16.37)		26.49 (21.53 - 32.13)	
1.00	3.36 (3.10 - 3.64)		4.77 (4.50 - 5.05)		4.57 (4.29 - 4.86)		5.20 (4.63 - 5.82)	
2.00	1.83 (1.65 - 2.03)		4.70 (4.44 - 4.99)		4.49 (4.23 - 4.78)		4.07 (3.82 - 4.35)	
5.00	1.07 (0.94 - 1.22)		4.53 (4.27 - 4.80)		4.35 (4.08 - 4.62)		3.96 (3.70 - 4.23)	
10.00	0.75 (0.63 - 0.89)		4.33 (4.08 - 4.59)		4.15 (3.89 - 4.43)		3.79 (3.54 - 4.05)	
20.00	0.50 (0.41 - 0.61)		3.87 (3.64 - 4.12)		3.76 (3.50 - 4.03)		3.40 (3.16 - 3.65)	
50.00	0.20 (0.15 - 0.27)		2.44 (2.26 - 2.64)		2.41 (2.19 - 2.65)		2.17 (1.98 - 2.38)	
Spam Misc%	Ham Misc%							
0.01	98.81 (95.00 - 99.72)		99.93 (99.62 - 99.99)		99.93 (99.68 - 99.98)		99.93 (99.55 - 99.99)	
0.02	96.21 (85.95 - 99.06)		99.78 (99.32 - 99.93)		99.78 (99.43 - 99.91)		99.78 (99.30 - 99.93)	
0.05	84.07 (66.07 - 93.46)		99.56 (99.11 - 99.78)		99.56 (99.25 - 99.74)		99.53 (99.12 - 99.75)	
0.10	72.82 (59.69 - 82.89)		98.30 (97.23 - 98.96)		98.09 (97.22 - 98.69)		98.36 (97.41 - 98.96)	
0.20	49.91 (39.82 - 60.01)		96.29 (94.82 - 97.36)		96.30 (95.09 - 97.22)		96.27 (94.83 - 97.32)	
0.50	20.28 (14.48 - 27.66)		90.64 (88.43 - 92.47)		90.65 (88.65 - 92.33)		89.60 (87.58 - 91.33)	
1.00	5.58 (4.31 - 7.19)		79.94 (77.04 - 82.55)		80.25 (77.26 - 82.93)		77.53 (74.46 - 80.34)	
2.00	1.78 (1.58 - 2.00)		59.32 (55.54 - 63.00)		59.04 (54.61 - 63.32)		55.42 (50.49 - 60.24)	
5.00	0.66 (0.62 - 0.70)		0.83 (0.58 - 1.19)		0.93 (0.87 - 0.98)		1.02 (0.98 - 1.07)	
10.00	0.54 (0.51 - 0.58)		0.58 (0.55 - 0.62)		0.50 (0.47 - 0.54)		0.60 (0.57 - 0.64)	
20.00	0.38 (0.35 - 0.41)		0.52 (0.49 - 0.56)		0.45 (0.42 - 0.49)		0.54 (0.51 - 0.58)	
50.00	0.16 (0.14 - 0.18)		0.33 (0.30 - 0.36)		0.28 (0.26 - 0.32)		0.33 (0.30 - 0.36)	



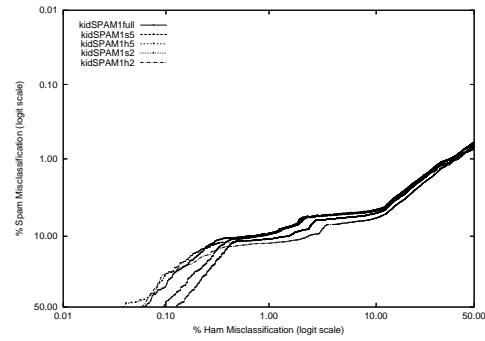
Learning Curves



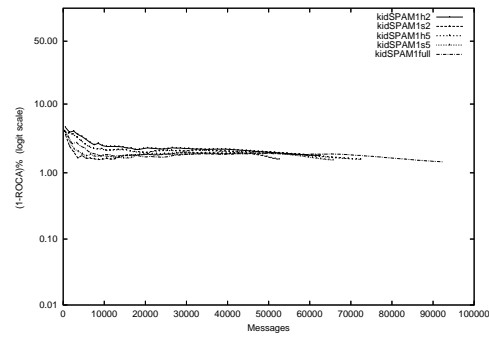
kid - Beijing U. of Posts and Telecom. - Public subset Corpora

	kidSPAM1full	kidSPAM1h2	kidSPAM1h5	kidSPAM1s2	kidSPAM1s5
Ham Misc%	0.91 (0.81-1.00)	0.40 (0.28-0.55)	0.36 (0.28-0.45)	1.99 (1.85-2.13)	1.44 (1.33-1.56)
Spam Misc%	9.40 (9.15-9.65)	13.24 (12.95-13.53)	12.01 (11.73-12.29)	6.74 (6.32-7.18)	8.01 (7.68-8.34)
Logit av Misc%	2.99 (2.83-3.15)	2.42 (2.07-2.82)	2.16 (1.93-2.43)	3.69 (3.52-3.86)	3.45 (3.29-3.61)
(1-ROCA)%	1.463 (1.399 - 1.529)	1.812 (1.721 - 1.907)	1.586 (1.514 - 1.662)	1.589 (1.463 - 1.726)	1.546 (1.459 - 1.639)

ROC Curves

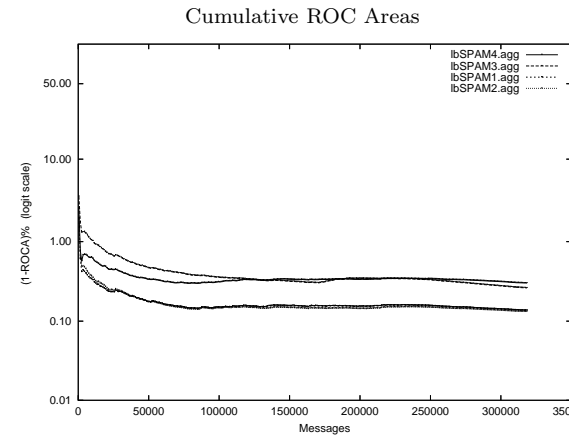
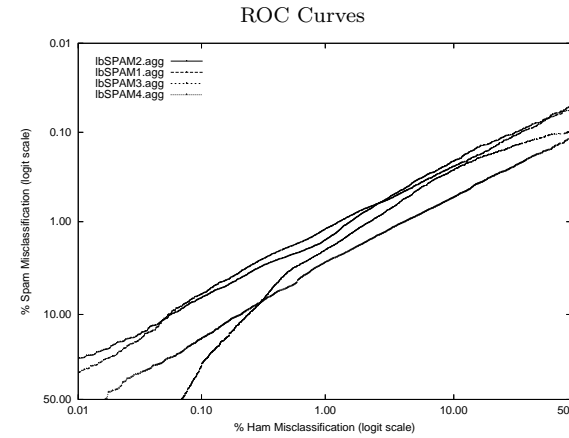


Cumulative ROC Area

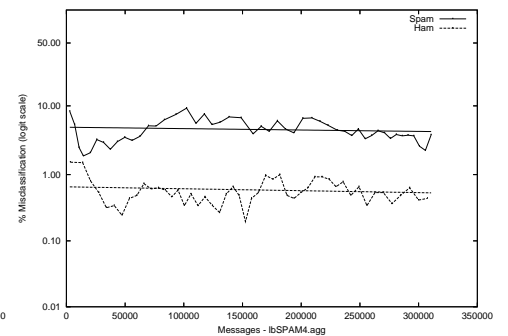
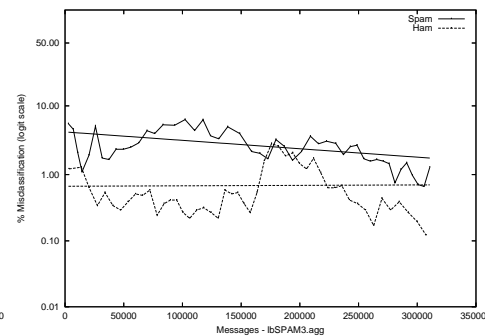
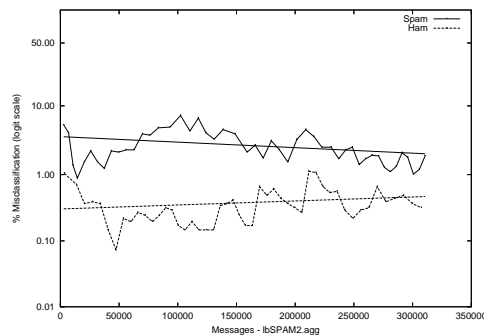
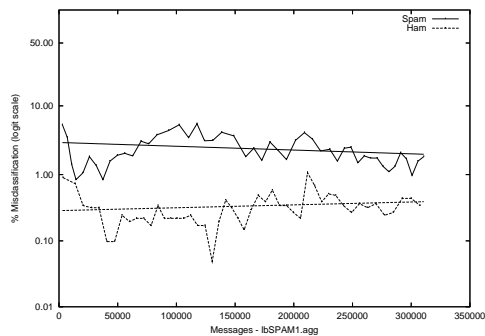


lb - Laird Breyer on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	lbSPAM1.agg		lbSPAM2.agg		lbSPAM3.agg		lbSPAM4.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.34 (0.31-0.36)		0.38 (0.35-0.41)		0.68 (0.65-0.72)		0.59 (0.56-0.62)	
Spam Misc%	2.46 (2.37-2.55)		2.75 (2.66-2.85)		2.81 (2.71-2.91)		4.66 (4.53-4.78)	
Logit av Misc%	0.91 (0.88-0.95)		1.03 (0.99-1.07)		1.39 (1.35-1.44)		1.67 (1.62-1.73)	
(1-ROCA)%	0.136 (0.125 - 0.148)		0.132 (0.121 - 0.144)		0.262 (0.244 - 0.281)		0.302 (0.284 - 0.321)	
Ham Misc%	Spam Misc%							
0.01	33.87 (27.94 - 40.36)		25.84 (19.78 - 32.99)		93.32 (92.21 - 94.29)		60.66 (49.19 - 71.07)	
0.02	24.44 (20.36 - 29.04)		19.31 (16.48 - 22.49)		89.07 (83.90 - 92.72)		43.87 (35.48 - 52.63)	
0.05	11.01 (8.41 - 14.28)		11.28 (9.66 - 13.14)		70.84 (63.58 - 77.16)		27.40 (24.42 - 30.60)	
0.10	6.19 (5.68 - 6.74)		6.75 (6.22 - 7.32)		29.95 (23.51 - 37.29)		17.23 (15.56 - 19.03)	
0.20	3.71 (3.39 - 4.06)		4.25 (3.98 - 4.54)		12.69 (11.22 - 14.32)		10.35 (9.65 - 11.09)	
0.50	1.94 (1.84 - 2.04)		2.41 (2.29 - 2.53)		3.59 (3.34 - 3.86)		5.24 (4.96 - 5.55)	
1.00	1.21 (1.14 - 1.29)		1.60 (1.51 - 1.70)		2.07 (1.97 - 2.19)		2.82 (2.67 - 2.98)	
2.00	0.74 (0.69 - 0.79)		0.80 (0.75 - 0.86)		1.12 (1.05 - 1.20)		1.76 (1.67 - 1.86)	
5.00	0.41 (0.37 - 0.45)		0.36 (0.33 - 0.40)		0.48 (0.44 - 0.53)		0.90 (0.85 - 0.96)	
10.00	0.24 (0.22 - 0.27)		0.21 (0.19 - 0.24)		0.27 (0.24 - 0.30)		0.54 (0.50 - 0.58)	
20.00	0.14 (0.12 - 0.16)		0.12 (0.11 - 0.15)		0.17 (0.15 - 0.19)		0.30 (0.26 - 0.33)	
50.00	0.06 (0.04 - 0.07)		0.05 (0.04 - 0.07)		0.10 (0.08 - 0.12)		0.12 (0.10 - 0.14)	
Spam Misc%	Ham Misc%							
0.01	91.25 (83.73 - 95.48)		89.16 (81.72 - 93.81)		99.64 (98.41 - 99.92)		93.46 (85.05 - 97.29)	
0.02	81.12 (74.67 - 86.22)		80.23 (68.09 - 88.53)		97.81 (90.17 - 99.54)		87.18 (83.13 - 90.37)	
0.05	55.67 (44.56 - 66.25)		51.66 (43.52 - 59.71)		85.51 (77.07 - 91.20)		77.35 (73.02 - 81.17)	
0.10	28.29 (23.23 - 33.97)		26.03 (20.65 - 32.23)		48.44 (34.26 - 62.88)		56.14 (49.25 - 62.81)	
0.20	13.19 (11.02 - 15.71)		10.91 (9.37 - 12.67)		14.59 (12.21 - 17.33)		30.70 (26.91 - 34.77)	
0.50	3.80 (3.39 - 4.26)		3.50 (3.18 - 3.86)		4.86 (4.49 - 5.27)		10.92 (9.84 - 12.09)	
1.00	1.30 (1.20 - 1.42)		1.60 (1.50 - 1.71)		2.31 (2.11 - 2.52)		4.31 (3.96 - 4.69)	
2.00	0.47 (0.43 - 0.52)		0.72 (0.66 - 0.79)		1.04 (0.97 - 1.11)		1.67 (1.55 - 1.80)	
5.00	0.13 (0.11 - 0.15)		0.16 (0.15 - 0.18)		0.39 (0.37 - 0.42)		0.55 (0.50 - 0.59)	
10.00	0.05 (0.04 - 0.07)		0.06 (0.05 - 0.07)		0.24 (0.22 - 0.27)		0.21 (0.19 - 0.23)	
20.00	0.03 (0.02 - 0.03)		0.02 (0.01 - 0.03)		0.14 (0.12 - 0.16)		0.08 (0.07 - 0.10)	
50.00	0.01 (0.00 - 0.01)		0.00 (0.00 - 0.01)		0.07 (0.06 - 0.08)		0.02 (0.01 - 0.02)	

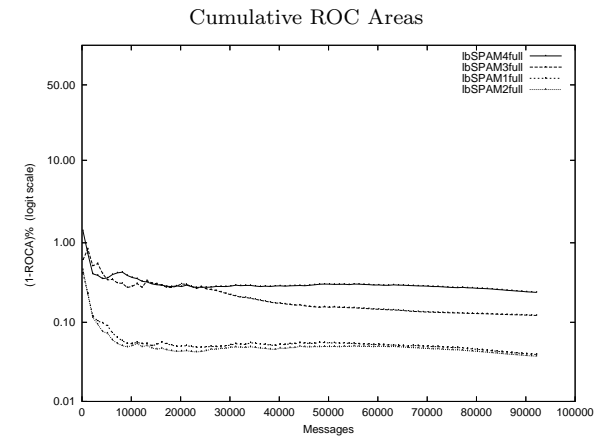
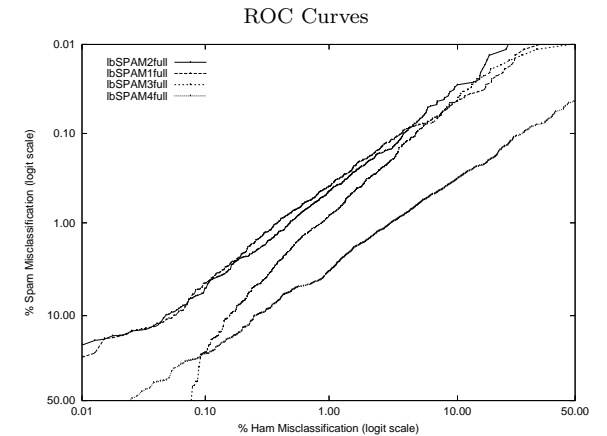


Learning Curves

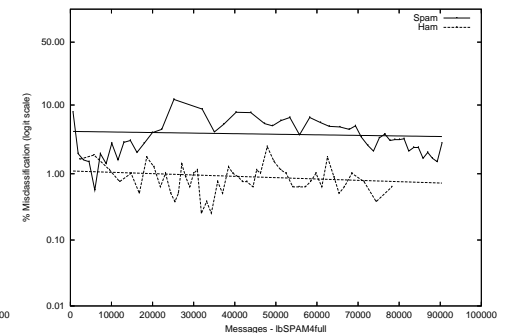
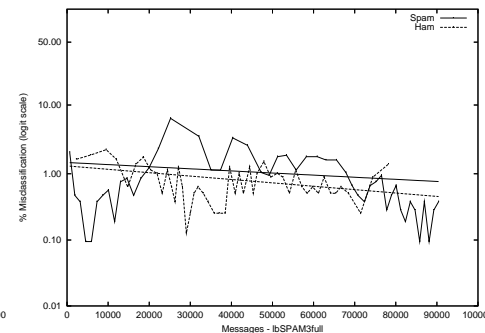
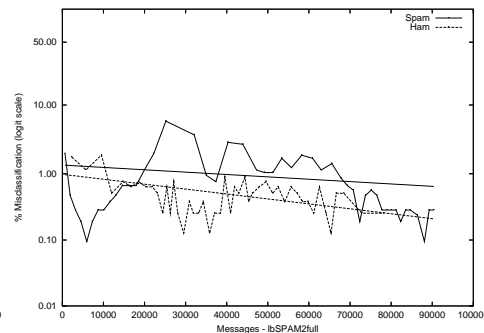
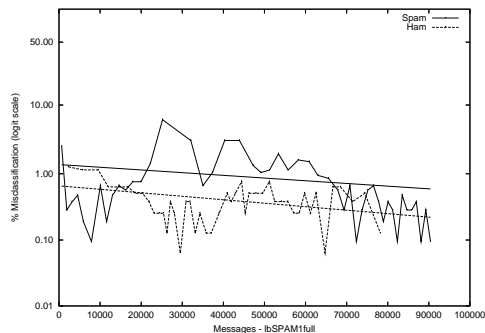


lb - Laird Breyer on Public trec05p-1/full Corpus

Gold Std. Filter Classification	lbSPAM1full		lbSPAM2full		lbSPAM3full		lbSPAM4full	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	39237 162	475 52315	39197 202	490 52300	39072 327	556 52234	39039 360	2041 50749
Ham Misc%	0.41 (0.35-0.48)		0.51 (0.44-0.59)		0.83 (0.74-0.92)		0.91 (0.82-1.01)	
Spam Misc%	0.90 (0.82-0.98)		0.93 (0.85-1.01)		1.05 (0.97-1.14)		3.87 (3.70-4.03)	
Logit av Misc%	0.61 (0.56-0.67)		0.69 (0.64-0.75)		0.94 (0.87-1.00)		1.89 (1.79-2.00)	
(1-ROCA)%	0.039 (0.031 - 0.049)		0.037 (0.031 - 0.045)		0.122 (0.098 - 0.153)		0.238 (0.217 - 0.261)	
Ham Misc%	Spam Misc%							
0.01	27.58 (17.88 - 39.97)		19.77 (13.39 - 28.22)		88.49 (82.73 - 92.51)		84.57 (55.09 - 96.07)	
0.02	16.44 (10.41 - 24.98)		15.68 (12.29 - 19.79)		87.61 (83.71 - 90.68)		64.28 (36.81 - 84.75)	
0.05	10.96 (7.81 - 15.15)		9.84 (6.64 - 14.32)		75.22 (57.72 - 87.10)		36.92 (27.28 - 47.73)	
0.10	4.56 (3.14 - 6.60)		5.19 (3.53 - 7.58)		22.38 (8.12 - 48.49)		22.94 (19.02 - 27.40)	
0.20	2.21 (1.62 - 3.01)		2.42 (1.98 - 2.96)		7.28 (4.95 - 10.58)		14.12 (11.33 - 17.47)	
0.50	0.77 (0.65 - 0.90)		0.96 (0.78 - 1.17)		1.91 (1.53 - 2.38)		5.50 (4.89 - 6.18)	
1.00	0.39 (0.33 - 0.47)		0.45 (0.37 - 0.54)		0.86 (0.75 - 0.98)		3.41 (2.98 - 3.90)	
2.00	0.19 (0.16 - 0.24)		0.22 (0.18 - 0.26)		0.33 (0.28 - 0.39)		1.55 (1.40 - 1.71)	
5.00	0.08 (0.06 - 0.11)		0.07 (0.05 - 0.10)		0.09 (0.07 - 0.13)		0.63 (0.56 - 0.71)	
10.00	0.04 (0.03 - 0.07)		0.03 (0.02 - 0.05)		0.04 (0.02 - 0.06)		0.32 (0.27 - 0.38)	
20.00	0.02 (0.01 - 0.04)		0.01 (0.01 - 0.03)		0.02 (0.01 - 0.04)		0.15 (0.12 - 0.19)	
50.00	0.01 (0.00 - 0.03)		0.00 (0.00 - 0.01)		0.01 (0.00 - 0.03)		0.04 (0.03 - 0.07)	
Spam Misc%	Ham Misc%							
0.01	36.72 (19.40 - 58.31)		22.51 (15.12 - 32.15)		54.47 (0.34 - 99.76)		81.45 (72.68 - 87.88)	
0.02	21.10 (13.23 - 31.95)		15.04 (10.42 - 21.21)		17.02 (6.39 - 38.11)		78.14 (70.01 - 84.56)	
0.05	8.68 (5.68 - 13.03)		6.28 (4.46 - 8.79)		8.76 (6.61 - 11.51)		45.38 (31.86 - 59.62)	
0.10	3.79 (2.66 - 5.36)		3.93 (3.27 - 4.72)		4.87 (3.92 - 6.05)		27.45 (20.85 - 35.21)	
0.20	1.86 (1.43 - 2.41)		2.15 (1.78 - 2.60)		3.20 (2.78 - 3.67)		16.73 (14.00 - 19.86)	
0.50	0.79 (0.68 - 0.91)		0.94 (0.81 - 1.09)		1.45 (1.29 - 1.63)		6.25 (5.44 - 7.16)	
1.00	0.38 (0.32 - 0.45)		0.49 (0.42 - 0.56)		0.87 (0.78 - 0.98)		3.11 (2.85 - 3.40)	
2.00	0.22 (0.17 - 0.27)		0.22 (0.17 - 0.32)		0.49 (0.42 - 0.56)		1.53 (1.39 - 1.68)	
5.00	0.10 (0.07 - 0.13)		0.10 (0.07 - 0.15)		0.25 (0.20 - 0.32)		0.57 (0.47 - 0.68)	
10.00	0.06 (0.04 - 0.08)		0.05 (0.03 - 0.08)		0.16 (0.12 - 0.21)		0.26 (0.21 - 0.32)	
20.00	0.02 (0.01 - 0.04)		0.01 (0.00 - 0.03)		0.11 (0.08 - 0.15)		0.12 (0.09 - 0.16)	
50.00	0.00 (0.00 - 0.00)		0.00 (0.00 - 0.00)		0.08 (0.05 - 0.12)		0.03 (0.01 - 0.05)	

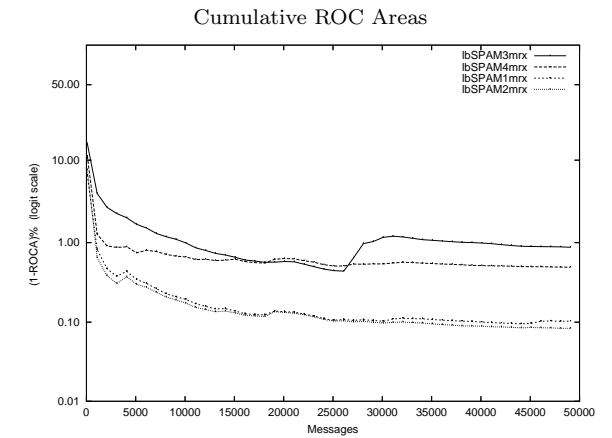
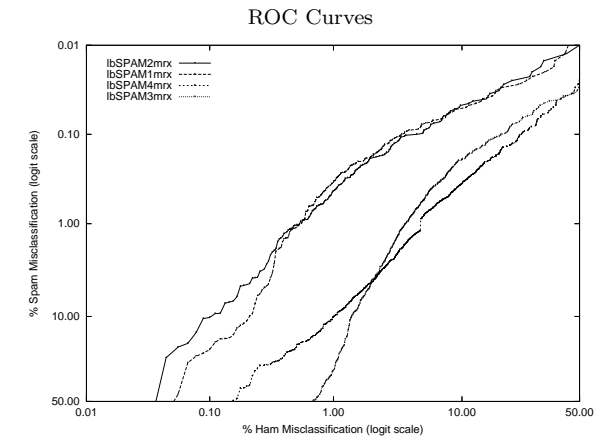


Learning Curves

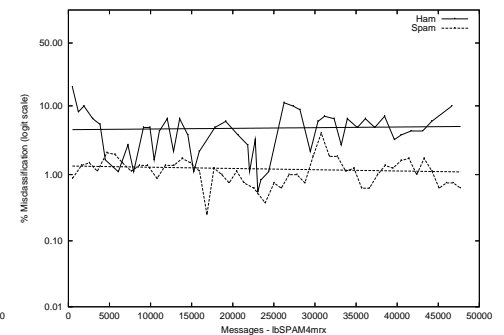
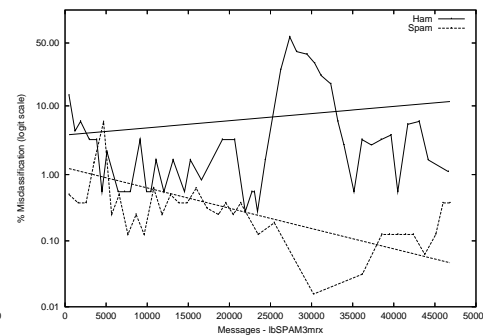
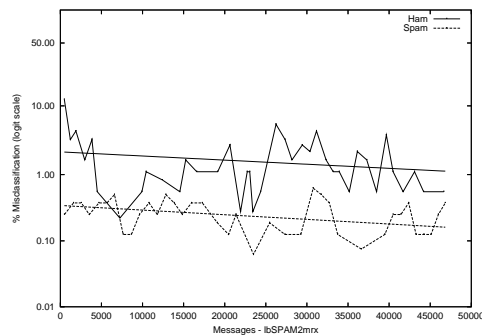
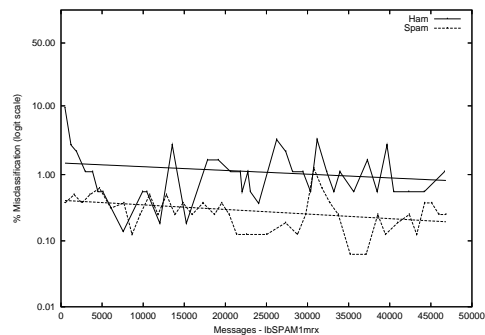


lb - Laird Breyer on Private Mr X. Corpus

Gold Std. Filter Classification	lbSPAM1mrx		lbSPAM2mrx		lbSPAM3mrx		lbSPAM4mrx	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	8935 103	113 39935	8891 147	94 39954	8420 618	139 39909	8599 439	485 39563
Ham Misc%	1.14 (0.93-1.38)		1.63 (1.38-1.91)		6.84 (6.33-7.38)		4.86 (4.42-5.32)	
Spam Misc%	0.28 (0.23-0.34)		0.23 (0.19-0.29)		0.35 (0.29-0.41)		1.21 (1.11-1.32)	
Logit av Misc%	0.57 (0.50-0.65)		0.62 (0.54-0.71)		1.57 (1.44-1.72)		2.44 (2.29-2.60)	
(1-ROCA)%	0.103 (0.072 - 0.148)		0.083 (0.054 - 0.130)		0.875 (0.723 - 1.059)		0.492 (0.428 - 0.566)	
Ham Misc%	Spam Misc%							
0.01	99.79 (74.15 - 100.00)		99.09 (78.96 - 99.97)		100.00 (nan - nan)		99.69 (91.29 - 99.99)	
0.02	69.58 (1.69 - 99.67)		80.96 (9.20 - 99.44)		99.48 (nan - nan)		90.49 (26.16 - 99.61)	
0.05	57.85 (25.64 - 84.53)		24.34 (1.90 - 84.26)		97.82 (89.99 - 99.55)		82.08 (62.23 - 92.72)	
0.10	20.67 (8.02 - 43.79)		10.24 (4.03 - 23.66)		95.73 (89.13 - 98.40)		58.36 (39.22 - 75.28)	
0.20	11.13 (5.35 - 21.73)		4.74 (2.58 - 8.56)		85.79 (70.42 - 93.87)		40.38 (26.66 - 55.78)	
0.50	1.10 (0.67 - 1.81)		1.03 (0.75 - 1.42)		66.66 (55.15 - 76.48)		20.06 (16.02 - 24.81)	
1.00	0.34 (0.24 - 0.49)		0.43 (0.30 - 0.62)		30.89 (18.44 - 46.91)		10.24 (7.93 - 13.13)	
2.00	0.18 (0.14 - 0.23)		0.19 (0.14 - 0.24)		4.30 (2.95 - 6.24)		4.50 (3.78 - 5.34)	
5.00	0.08 (0.05 - 0.11)		0.10 (0.07 - 0.14)		0.58 (0.48 - 0.69)		0.87 (0.61 - 1.25)	
10.00	0.05 (0.03 - 0.08)		0.05 (0.03 - 0.08)		0.19 (0.15 - 0.24)		0.36 (0.30 - 0.44)	
20.00	0.03 (0.02 - 0.06)		0.03 (0.01 - 0.05)		0.10 (0.07 - 0.13)		0.14 (0.11 - 0.19)	
50.00	0.01 (0.00 - 0.03)		0.01 (0.00 - 0.03)		0.03 (0.02 - 0.06)		0.02 (0.01 - 0.05)	
Spam Misc%	Ham Misc%							
0.01	44.80 (19.66 - 72.91)		49.81 (28.83 - 70.86)		96.05 (67.63 - 99.65)		83.74 (22.24 - 98.93)	
0.02	34.34 (16.98 - 57.22)		29.04 (14.19 - 50.33)		63.34 (23.18 - 90.82)		52.56 (22.77 - 80.63)	
0.05	10.64 (5.51 - 19.56)		9.53 (4.67 - 18.46)		31.58 (19.58 - 46.67)		37.97 (29.87 - 46.81)	
0.10	3.45 (2.15 - 5.49)		4.54 (2.74 - 7.41)		18.72 (13.96 - 24.63)		25.18 (20.24 - 30.86)	
0.20	1.74 (1.24 - 2.43)		1.89 (1.37 - 2.62)		9.38 (7.79 - 11.26)		15.38 (12.54 - 18.72)	
0.50	0.77 (0.60 - 1.00)		0.93 (0.73 - 1.19)		5.34 (4.85 - 5.89)		7.94 (6.88 - 9.16)	
1.00	0.55 (0.42 - 0.73)		0.54 (0.40 - 0.73)		3.73 (3.31 - 4.20)		4.89 (4.50 - 5.31)	
2.00	0.37 (0.26 - 0.51)		0.33 (0.23 - 0.49)		2.82 (2.47 - 3.21)		3.33 (3.00 - 3.70)	
5.00	0.28 (0.20 - 0.39)		0.18 (0.11 - 0.30)		1.95 (1.66 - 2.28)		1.81 (1.58 - 2.08)	
10.00	0.21 (0.14 - 0.31)		0.11 (0.06 - 0.20)		1.41 (1.17 - 1.69)		1.02 (0.83 - 1.25)	
20.00	0.11 (0.06 - 0.20)		0.06 (0.02 - 0.14)		1.18 (0.96 - 1.46)		0.51 (0.39 - 0.66)	
50.00	0.06 (0.02 - 0.16)		0.04 (0.01 - 0.15)		0.69 (0.52 - 0.90)		0.15 (0.10 - 0.25)	

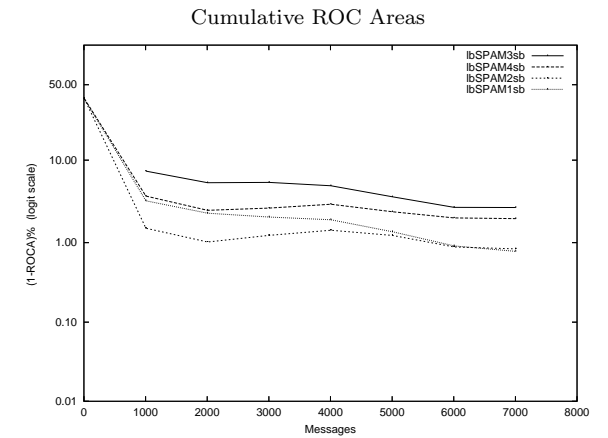
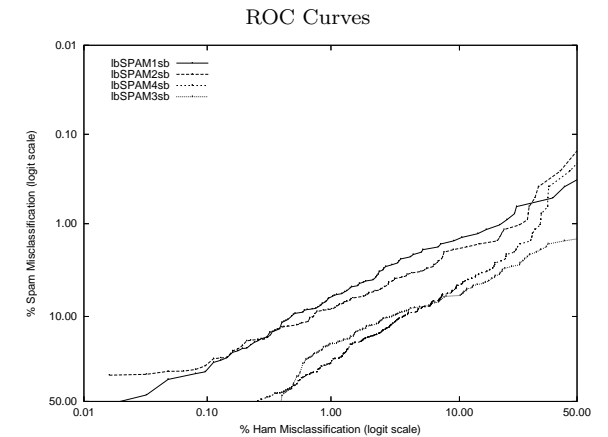


Learning Curves

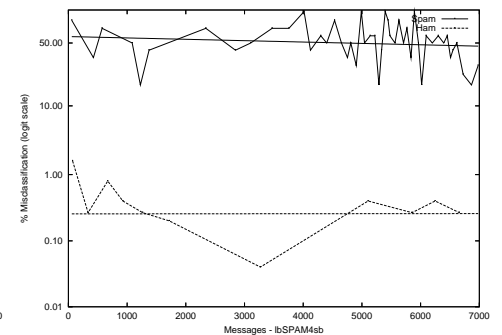
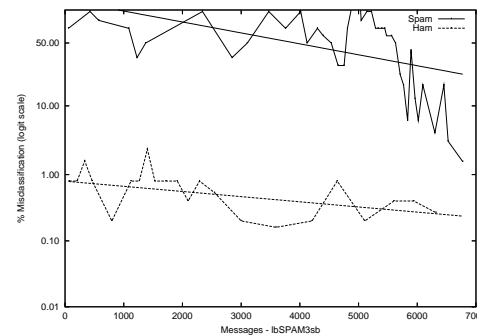
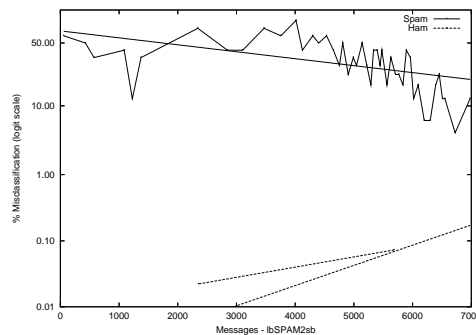
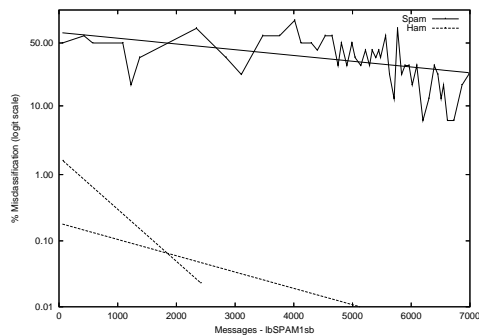


lb - Laird Breyer on Private S.B. Corpus

Gold Std. Filter Classification	lbSPAM1sb			lbSPAM2sb			lbSPAM3sb			lbSPAM4sb		
	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	spam ham	ham spam	ham spam	
Ham Misc%	0.05 (0.01-0.14)			0.03 (0.00-0.12)			0.47 (0.31-0.67)			0.26 (0.15-0.42)		
Spam Misc%	36.13 (32.74-39.62)			33.16 (29.85-36.60)			42.45 (38.94-46.02)			49.94 (46.36-53.51)		
Logit av Misc%	1.62 (0.92-2.84)			1.25 (0.62-2.47)			5.55 (4.61-6.67)			4.82 (3.78-6.14)		
(1-ROCA)%	0.778 (0.449 - 1.347)			0.835 (0.593 - 1.173)			2.727 (2.128 - 3.488)			1.988 (1.608 - 2.455)		
Ham Misc%										Spam Misc%		
0.01	82.19 (49.41 - 95.62)			75.61 (35.31 - 94.63)			98.84 (97.39 - 99.49)			71.10 (63.40 - 77.75)		
0.02	51.74 (20.32 - 81.84)			33.42 (8.48 - 73.11)			98.71 (97.33 - 99.38)			67.74 (57.59 - 76.46)		
0.05	36.13 (17.63 - 59.93)			31.35 (17.29 - 49.96)			98.58 (97.03 - 99.33)			60.90 (50.18 - 70.67)		
0.10	31.61 (20.07 - 45.98)			28.52 (20.45 - 38.24)			98.32 (96.53 - 99.20)			52.65 (44.06 - 61.08)		
0.20	20.13 (13.46 - 29.00)			19.10 (12.82 - 27.48)			97.29 (89.42 - 99.35)			50.58 (45.77 - 55.38)		
0.50	9.42 (6.53 - 13.41)			12.52 (9.94 - 15.64)			38.84 (14.28 - 70.77)			40.00 (32.25 - 48.28)		
1.00	6.45 (4.65 - 8.89)			8.39 (6.68 - 10.48)			18.45 (15.27 - 22.12)			27.23 (22.23 - 32.87)		
2.00	4.26 (3.10 - 5.82)			5.68 (4.21 - 7.62)			12.52 (10.14 - 15.35)			16.00 (13.44 - 18.94)		
5.00	2.19 (1.33 - 3.60)			3.48 (2.49 - 4.86)			8.13 (6.36 - 10.33)			8.39 (6.51 - 10.74)		
10.00	1.55 (0.83 - 2.87)			1.94 (1.20 - 3.11)			6.19 (4.80 - 7.95)			4.77 (3.41 - 6.64)		
20.00	1.03 (0.42 - 2.52)			1.29 (0.67 - 2.48)			3.23 (2.28 - 4.54)			2.58 (1.69 - 3.92)		
50.00	0.39 (0.09 - 1.66)			0.26 (0.06 - 1.06)			1.55 (0.87 - 2.73)			0.26 (0.07 - 0.94)		
Spam Misc%										Ham Misc%		
0.01	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.10	81.98 (58.00 - 93.74)			54.68 (44.89 - 64.12)			100.00 (nan - nan)			80.89 (53.22 - 94.03)		
0.20	79.52 (42.05 - 95.41)			52.58 (35.93 - 68.67)			99.61 (nan - nan)			57.90 (26.07 - 84.29)		
0.50	44.12 (13.77 - 79.61)			32.68 (18.13 - 51.54)			71.75 (nan - nan)			37.01 (24.40 - 51.67)		
1.00	21.14 (7.91 - 45.54)			28.33 (17.24 - 42.86)			60.79 (35.92 - 81.09)			33.65 (28.60 - 39.11)		
2.00	5.30 (1.79 - 14.61)			8.99 (3.89 - 19.40)			33.45 (17.91 - 53.64)			24.27 (17.12 - 33.19)		
5.00	1.44 (0.88 - 2.37)			2.66 (1.61 - 4.37)			12.45 (8.36 - 18.15)			9.45 (6.54 - 13.47)		
10.00	0.48 (0.30 - 0.77)			0.72 (0.48 - 1.08)			3.07 (2.01 - 4.65)			3.77 (2.66 - 5.33)		
20.00	0.22 (0.12 - 0.41)			0.19 (0.09 - 0.40)			0.87 (0.61 - 1.23)			1.43 (1.13 - 1.80)		
50.00	0.03 (0.01 - 0.14)			0.02 (0.00 - 0.06)			0.40 (0.25 - 0.64)			0.26 (0.08 - 0.80)		

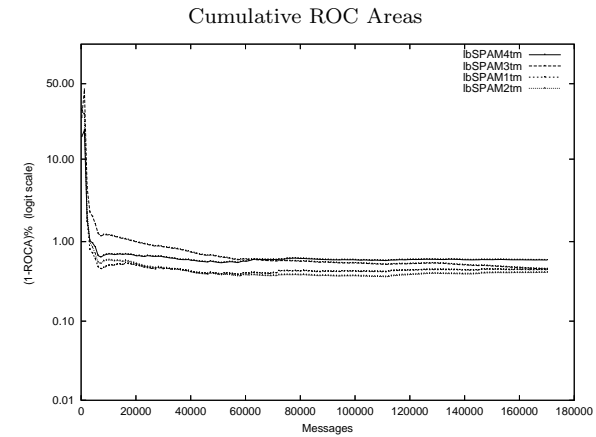
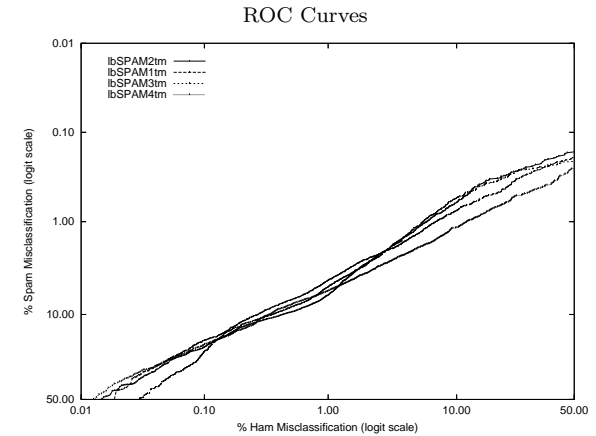


Learning Curves

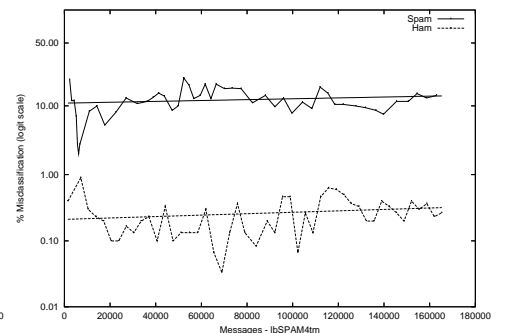
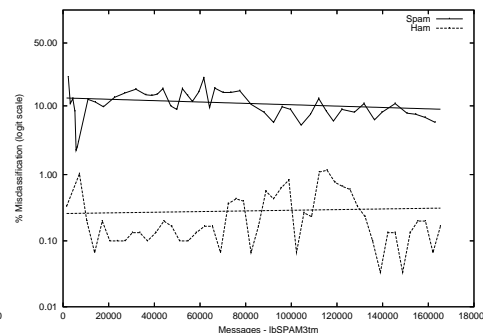
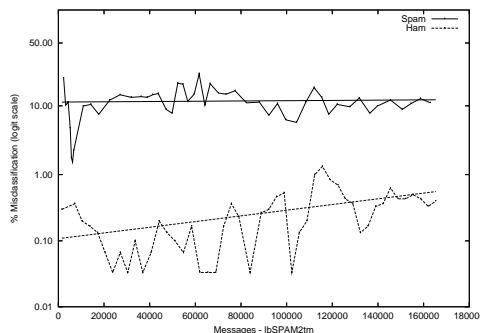
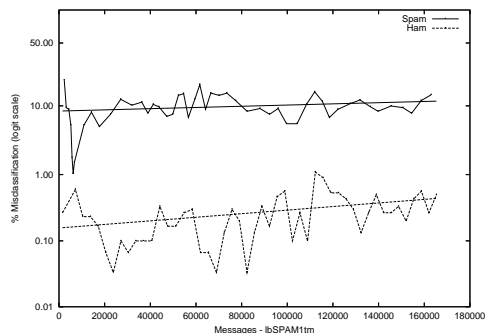


lb - Laird Breyer on Private T.M. Corpus

Gold Std. Filter Classification	lbSPAM1tm			lbSPAM2tm			lbSPAM3tm			lbSPAM4tm		
	ham spam	ham 150262 423	spam 1913 17603	ham spam	ham 150255 430	spam 2270 17246	ham spam	ham 150254 431	spam 2154 17362	ham spam	ham 150288 397	spam 2354 17162
Ham Misc%	0.28 (0.25-0.31)			0.29 (0.26-0.31)			0.29 (0.26-0.31)			0.26 (0.24-0.29)		
Spam Misc%	9.80 (9.39-10.23)			11.63 (11.18-12.09)			11.04 (10.60-11.49)			12.06 (11.61-12.53)		
Logit av Misc%	1.72 (1.63-1.81)			1.90 (1.81-2.00)			1.85 (1.76-1.95)			1.87 (1.77-1.97)		
(1-ROCA)%	0.443 (0.388 - 0.506)			0.411 (0.364 - 0.464)			0.456 (0.402 - 0.518)			0.588 (0.529 - 0.655)		
Ham Misc%	Spam Misc%											
0.01	61.04 (51.62 - 69.70)			56.63 (44.37 - 68.13)			74.45 (47.93 - 90.22)			54.03 (45.59 - 62.25)		
0.02	44.24 (33.82 - 55.19)			43.53 (36.98 - 50.31)			61.49 (50.30 - 71.58)			39.36 (33.71 - 45.30)		
0.05	27.17 (23.97 - 30.61)			28.42 (24.07 - 33.20)			37.17 (32.31 - 42.30)			26.81 (23.90 - 29.93)		
0.10	17.94 (16.57 - 19.39)			20.53 (18.71 - 22.48)			22.38 (18.71 - 26.52)			19.67 (18.16 - 21.28)		
0.20	11.96 (11.05 - 12.94)			13.39 (12.46 - 14.39)			12.76 (11.75 - 13.84)			14.15 (13.33 - 15.01)		
0.50	7.13 (6.78 - 7.50)			9.50 (9.03 - 9.99)			8.50 (8.04 - 8.98)			8.03 (7.58 - 8.50)		
1.00	4.41 (4.11 - 4.72)			6.32 (5.87 - 6.80)			5.21 (4.85 - 5.60)			5.72 (5.37 - 6.08)		
2.00	2.63 (2.40 - 2.89)			2.93 (2.65 - 3.24)			3.04 (2.79 - 3.32)			3.72 (3.44 - 4.02)		
5.00	1.40 (1.26 - 1.56)			1.20 (1.05 - 1.37)			1.15 (1.00 - 1.33)			2.00 (1.81 - 2.21)		
10.00	0.76 (0.65 - 0.89)			0.61 (0.52 - 0.73)			0.55 (0.45 - 0.66)			1.16 (1.03 - 1.32)		
20.00	0.47 (0.39 - 0.57)			0.33 (0.26 - 0.42)			0.35 (0.27 - 0.45)			0.67 (0.57 - 0.79)		
50.00	0.19 (0.14 - 0.27)			0.17 (0.12 - 0.24)			0.22 (0.16 - 0.29)			0.26 (0.19 - 0.34)		
Spam Misc%	Ham Misc%											
0.01	99.86 (95.84 - 100.00)			99.77 (97.87 - 99.98)			99.99 (nan - nan)			99.49 (95.66 - 99.94)		
0.02	98.33 (78.52 - 99.89)			98.11 (81.90 - 99.83)			99.96 (99.56 - 100.00)			99.17 (92.86 - 99.91)		
0.05	91.91 (74.62 - 97.77)			89.79 (58.42 - 98.21)			99.36 (97.56 - 99.83)			89.66 (73.58 - 96.43)		
0.10	80.38 (61.71 - 91.24)			66.68 (54.11 - 77.26)			97.05 (85.02 - 99.48)			79.15 (68.07 - 87.11)		
0.20	48.38 (28.46 - 68.83)			37.98 (24.08 - 54.17)			59.47 (30.76 - 82.89)			59.03 (47.37 - 69.75)		
0.50	17.90 (13.44 - 23.44)			11.99 (10.35 - 13.85)			11.36 (8.67 - 14.75)			27.52 (20.64 - 35.65)		
1.00	7.08 (6.01 - 8.31)			6.00 (5.19 - 6.93)			5.60 (4.96 - 6.32)			11.81 (10.08 - 13.80)		
2.00	3.18 (2.68 - 3.77)			3.04 (2.72 - 3.39)			3.03 (2.76 - 3.33)			5.00 (4.36 - 5.73)		
5.00	0.87 (0.80 - 0.95)			1.23 (1.16 - 1.31)			1.04 (0.95 - 1.13)			1.25 (1.11 - 1.41)		
10.00	0.27 (0.24 - 0.31)			0.44 (0.38 - 0.50)			0.36 (0.32 - 0.41)			0.34 (0.31 - 0.38)		
20.00	0.08 (0.07 - 0.10)			0.10 (0.09 - 0.12)			0.11 (0.10 - 0.13)			0.10 (0.08 - 0.12)		
50.00	0.02 (0.01 - 0.03)			0.01 (0.01 - 0.02)			0.03 (0.02 - 0.04)			0.01 (0.01 - 0.02)		



Learning Curves

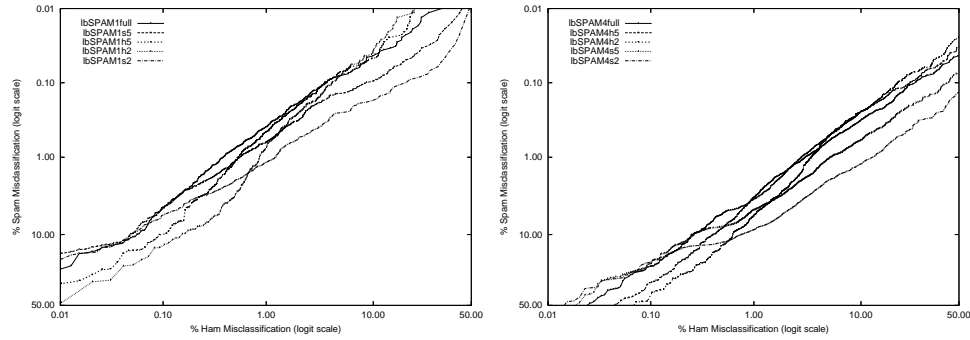


lb - Laird Breyer - Public subset Corpora

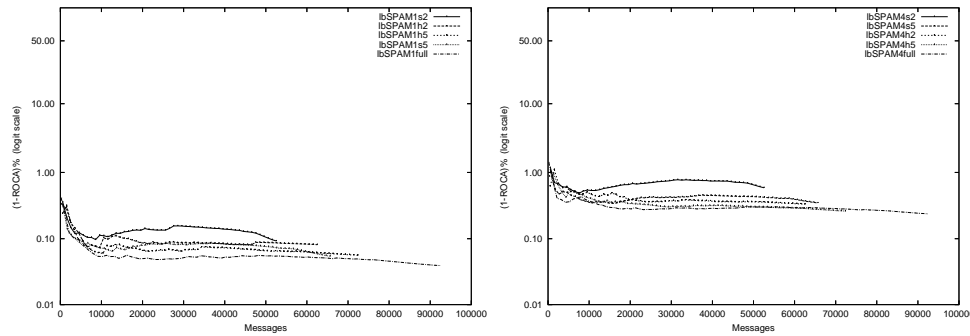
	lbSPAM1full	lbSPAM1h2	lbSPAM1h5	lbSPAM1s2	lbSPAM1s5
Ham Misc%	0.41 (0.35-0.48)	1.68 (1.44-1.96)	0.85 (0.73-0.99)	0.16 (0.12-0.20)	0.28 (0.23-0.34)
Spam Misc%	0.90 (0.82-0.98)	0.31 (0.27-0.36)	0.58 (0.52-0.65)	4.33 (3.98-4.69)	1.95 (1.78-2.12)
Logit av Misc%	0.61 (0.56-0.67)	0.73 (0.65-0.81)	0.71 (0.64-0.77)	0.84 (0.73-0.95)	0.74 (0.67-0.82)
(1-ROCA)%	0.039 (0.031 - 0.049)	0.081 (0.058 - 0.113)	0.056 (0.042 - 0.074)	0.092 (0.071 - 0.119)	0.054 (0.043 - 0.068)

	lbSPAM4full	lbSPAM4h2	lbSPAM4h5	lbSPAM4s2	lbSPAM4s5
Ham Misc%	0.91 (0.82-1.01)	2.96 (2.64-3.32)	1.44 (1.28-1.62)	0.58 (0.51-0.66)	0.71 (0.63-0.80)
Spam Misc%	3.87 (3.70-4.03)	1.46 (1.36-1.57)	2.57 (2.44-2.71)	11.69 (11.14-12.25)	6.94 (6.63-7.25)
Logit av Misc%	1.89 (1.79-2.00)	2.08 (1.95-2.23)	1.93 (1.81-2.05)	2.71 (2.53-2.90)	2.26 (2.13-2.41)
(1-ROCA)%	0.238 (0.217 - 0.261)	0.332 (0.284 - 0.388)	0.261 (0.229 - 0.298)	0.588 (0.523 - 0.661)	0.347 (0.318 - 0.378)

ROC Curves

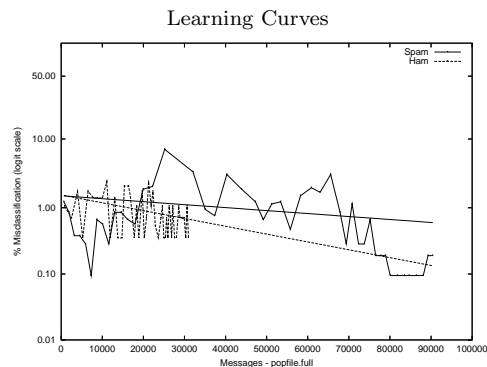
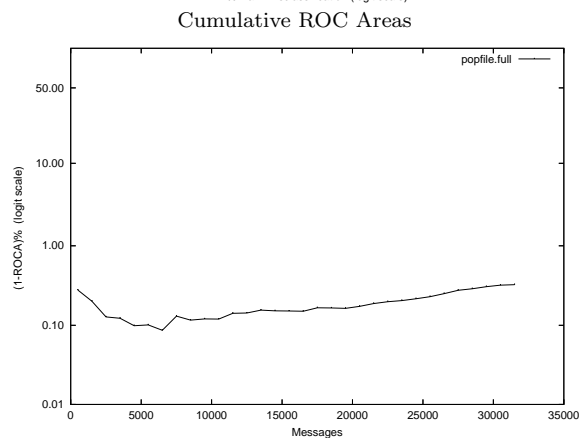
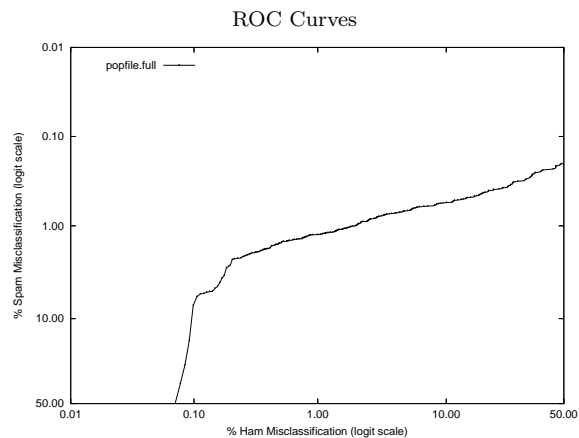


Cumulative ROC Area



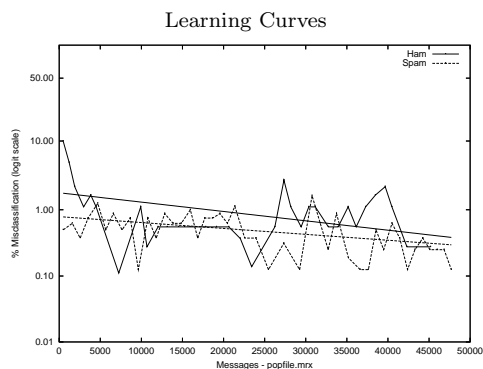
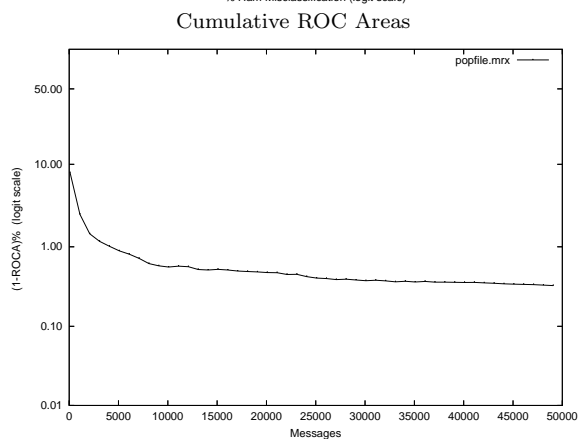
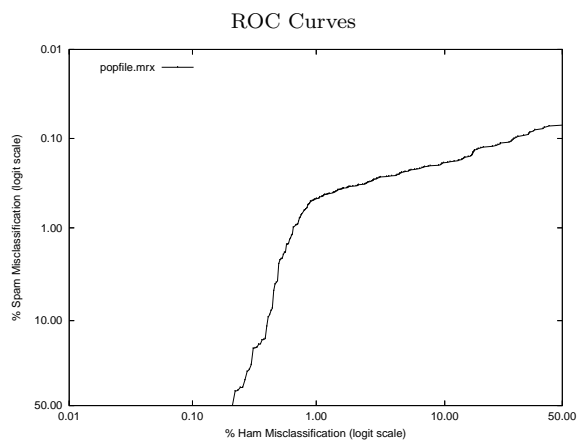
popfile - Popfile (non-participant) on Public trec05p-1/full Corpus

Gold Std.	popfile.full	
	ham	spam
Filter	14100	217
Classification	spam	131 17056
Ham Misc%	0.92 (0.77-1.09)	
Spam Misc%	1.26 (1.10-1.43)	
Logit av Misc%	0.94 (0.85-1.03)	
(1-ROCA)%	0.325 (0.275 - 0.384)	
Ham Misc%	Spam Misc%	
0.01	92.35 (72.08 - 98.26)	
0.02	86.48 (67.33 - 95.21)	
0.05	78.74 (29.18 - 97.08)	
0.10	7.35 (0.60 - 50.99)	
0.20	2.72 (1.51 - 4.85)	
0.50	1.56 (1.35 - 1.80)	
1.00	1.25 (1.11 - 1.40)	
2.00	1.00 (0.87 - 1.16)	
5.00	0.68 (0.58 - 0.81)	
10.00	0.56 (0.46 - 0.67)	
20.00	0.41 (0.32 - 0.51)	
50.00	0.20 (0.15 - 0.28)	
Spam Misc%	Ham Misc%	
0.01	91.33 (68.75 - 98.06)	
0.02	88.24 (77.91 - 94.11)	
0.05	76.68 (68.59 - 83.21)	
0.10	66.29 (57.81 - 73.84)	
0.20	53.80 (40.72 - 66.37)	
0.50	13.80 (7.84 - 23.16)	
1.00	2.01 (1.54 - 2.62)	
2.00	0.30 (0.19 - 0.45)	
5.00	0.15 (0.09 - 0.24)	
10.00	0.10 (0.06 - 0.17)	
20.00	0.09 (0.05 - 0.16)	
50.00	0.08 (0.04 - 0.14)	



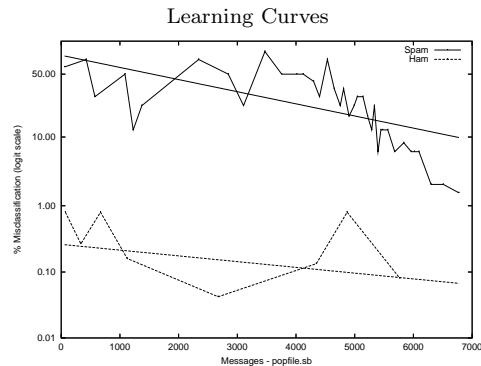
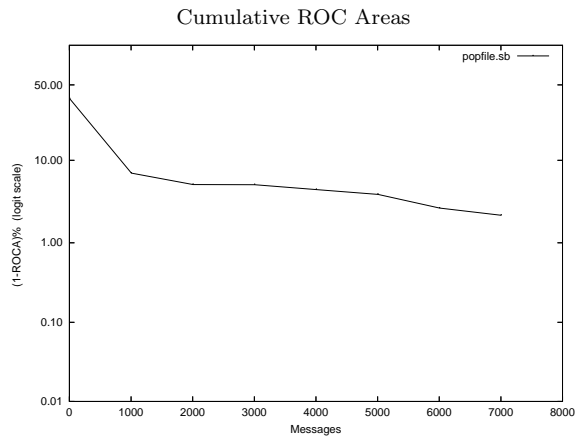
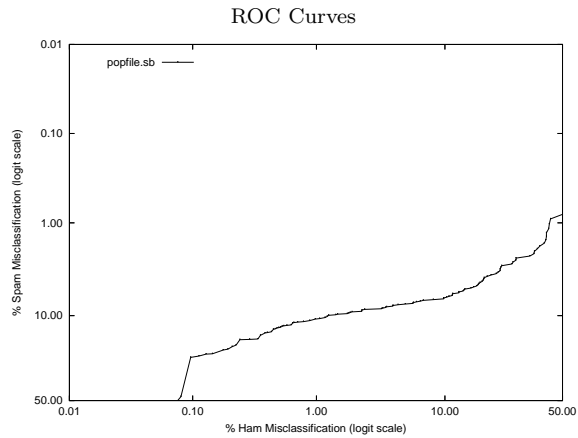
popfile - Popfile (non-participant) on Private Mr X. Corpus

Gold Std.	popfile.mrx	
	ham	spam
Filter	8951	197
Classification	spam	87 39851
Ham Misc%	0.96 (0.77-1.19)	
Spam Misc%	0.49 (0.43-0.57)	
Logit av Misc%	0.69 (0.61-0.78)	
(1-ROCA)%	0.326 (0.249 - 0.427)	
Ham Misc%	Spam Misc%	
0.01	99.57 (96.97 - 99.94)	
0.02	96.63 (77.52 - 99.58)	
0.05	93.84 (82.51 - 98.01)	
0.10	86.94 (75.20 - 93.60)	
0.20	67.46 (28.34 - 91.58)	
0.50	2.50 (0.59 - 9.88)	
1.00	0.47 (0.38 - 0.59)	
2.00	0.34 (0.29 - 0.41)	
5.00	0.24 (0.19 - 0.29)	
10.00	0.19 (0.15 - 0.24)	
20.00	0.12 (0.09 - 0.17)	
50.00	0.07 (0.05 - 0.11)	
Spam Misc%	Ham Misc%	
0.01	94.87 (81.67 - 98.71)	
0.02	87.28 (68.63 - 95.56)	
0.05	63.22 (47.46 - 76.59)	
0.10	28.87 (16.37 - 45.69)	
0.20	9.02 (4.67 - 16.71)	
0.50	0.93 (0.70 - 1.24)	
1.00	0.65 (0.51 - 0.83)	
2.00	0.54 (0.42 - 0.70)	
5.00	0.45 (0.34 - 0.60)	
10.00	0.41 (0.30 - 0.55)	
20.00	0.31 (0.22 - 0.44)	
50.00	0.22 (0.14 - 0.34)	



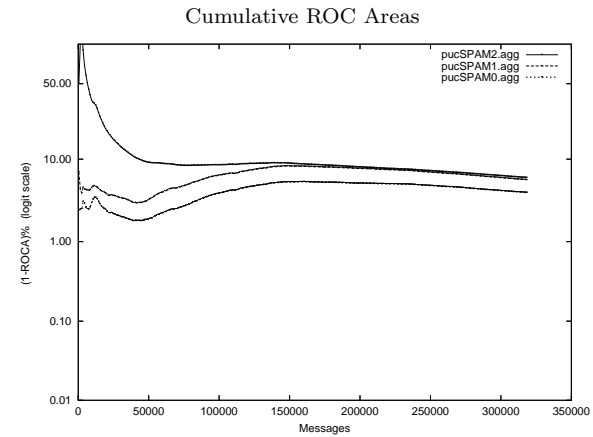
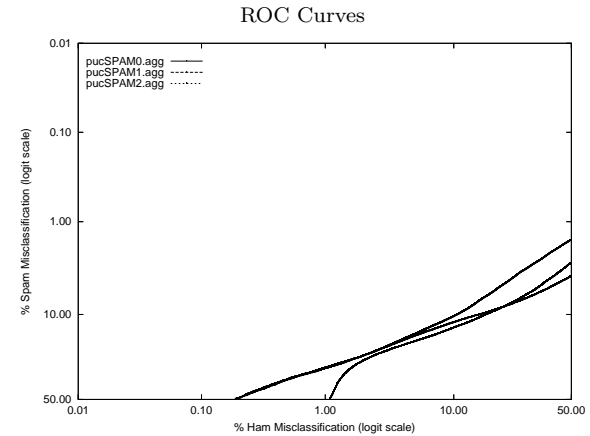
popfile - Popfile (non-participant) on Private S.B. Corpus

Gold Std.	popfile.sb	
	ham	spam
Filter	ham	6222 178
Classification	spam	9 597
Ham Misc%	0.14 (0.07-0.27)	
Spam Misc%	22.97 (20.05-26.09)	
Logit av Misc%	2.03 (1.46-2.83)	
(1-ROCA)%	2.199 (1.648 - 2.928)	
Ham Misc%	Spam Misc%	
0.01	93.29 (74.90 - 98.48)	
0.02	92.77 (63.46 - 98.96)	
0.05	59.35 (13.66 - 93.10)	
0.10	24.65 (7.45 - 57.06)	
0.20	20.90 (13.90 - 30.20)	
0.50	13.16 (10.36 - 16.58)	
1.00	10.84 (8.66 - 13.49)	
2.00	9.16 (7.12 - 11.71)	
5.00	7.74 (5.88 - 10.13)	
10.00	6.58 (4.92 - 8.75)	
20.00	3.87 (2.62 - 5.68)	
50.00	0.90 (0.40 - 2.00)	
Spam Misc%	Ham Misc%	
0.01	100.00 (nan - nan)	
0.02	100.00 (nan - nan)	
0.05	100.00 (nan - nan)	
0.10	83.68 (66.03 - 93.11)	
0.20	70.42 (48.25 - 85.88)	
0.50	58.58 (43.48 - 72.22)	
1.00	44.42 (34.66 - 54.64)	
2.00	38.44 (27.77 - 50.35)	
5.00	16.75 (10.68 - 25.30)	
10.00	1.25 (0.46 - 3.33)	
20.00	0.21 (0.11 - 0.39)	
50.00	0.08 (0.03 - 0.25)	

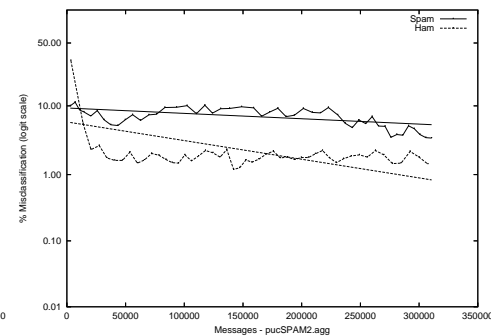
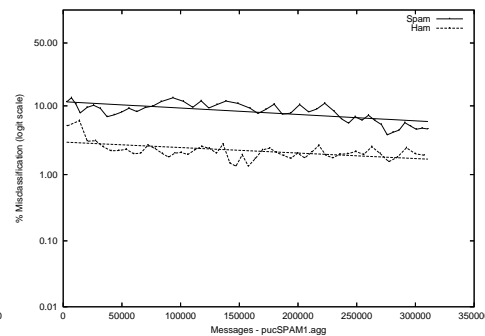
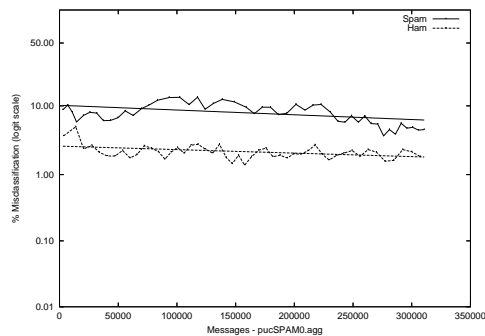


puc - PUCRS on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	pucSPAM0.agg		pucSPAM1.agg		pucSPAM2.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	2.21 (2.15-2.28)		2.30 (2.24-2.37)		2.58 (2.51-2.65)	
Spam Misc%	8.06 (7.91-8.22)		8.38 (8.22-8.54)		7.17 (7.02-7.32)	
Logit av Misc%	4.27 (4.19-4.34)		4.44 (4.36-4.51)		4.33 (4.25-4.40)	
(1-ROCA)%	4.030 (3.951 - 4.111)		5.746 (5.647 - 5.846)		6.107 (6.014 - 6.202)	
Ham Misc%	Spam Misc%					
0.01	81.31 (76.82 - 85.09)		80.07 (74.55 - 84.63)		99.99 (99.98 - 99.99)	
0.02	77.87 (75.46 - 80.11)		73.76 (72.53 - 74.96)		99.98 (99.98 - 99.99)	
0.05	68.08 (65.87 - 70.22)		65.09 (62.33 - 67.75)		99.98 (99.97 - 99.99)	
0.10	59.56 (57.81 - 61.29)		57.60 (55.87 - 59.30)		99.98 (99.97 - 99.99)	
0.20	49.51 (48.08 - 50.94)		49.15 (48.07 - 50.24)		99.98 (99.97 - 99.99)	
0.50	37.34 (36.43 - 38.26)		37.68 (37.06 - 38.30)		99.42 (98.93 - 99.69)	
1.00	30.62 (30.14 - 31.11)		31.07 (30.67 - 31.48)		57.54 (53.83 - 61.16)	
2.00	24.49 (24.08 - 24.91)		24.57 (24.23 - 24.92)		27.03 (26.54 - 27.52)	
5.00	15.78 (15.51 - 16.06)		16.37 (16.14 - 16.60)		18.45 (18.20 - 18.70)	
10.00	10.45 (10.25 - 10.65)		12.03 (11.84 - 12.23)		13.59 (13.36 - 13.82)	
20.00	5.44 (5.29 - 5.59)		8.86 (8.69 - 9.03)		9.01 (8.82 - 9.20)	
50.00	1.59 (1.52 - 1.67)		3.95 (3.83 - 4.08)		2.85 (2.76 - 2.95)	
Spam Misc%	Ham Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (99.99 - 100.00)		100.00 (99.99 - 100.00)		100.00 (99.99 - 100.00)	
0.05	99.99 (99.99 - 100.00)		99.99 (99.99 - 100.00)		99.99 (99.99 - 100.00)	
0.10	99.99 (99.98 - 99.99)		99.99 (99.98 - 99.99)		99.99 (99.98 - 99.99)	
0.20	99.97 (99.96 - 99.97)		99.97 (99.96 - 99.97)		99.97 (99.96 - 99.97)	
0.50	99.94 (99.91 - 99.96)		99.94 (99.91 - 99.95)		99.93 (99.88 - 99.96)	
1.00	68.64 (64.95 - 72.10)		84.94 (83.25 - 86.49)		80.04 (77.40 - 82.44)	
2.00	43.25 (42.01 - 44.50)		69.63 (68.56 - 70.68)		59.41 (58.39 - 60.42)	
5.00	21.53 (21.00 - 22.07)		41.87 (40.88 - 42.87)		36.14 (35.48 - 36.81)	
10.00	10.64 (10.34 - 10.96)		15.48 (14.82 - 16.16)		17.07 (16.59 - 17.56)	
20.00	3.18 (3.07 - 3.29)		3.26 (3.15 - 3.36)		4.06 (3.91 - 4.20)	
50.00	0.19 (0.18 - 0.21)		0.19 (0.17 - 0.21)		1.09 (1.05 - 1.13)	

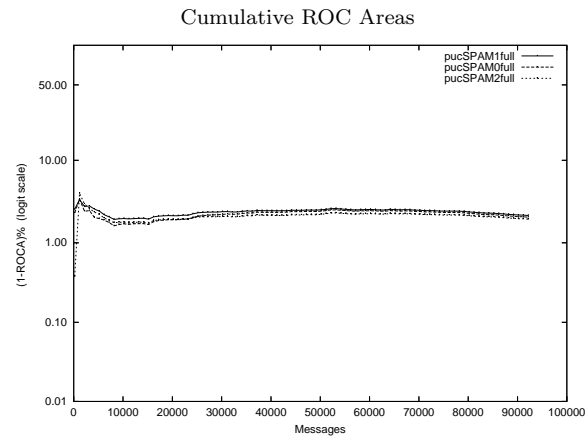
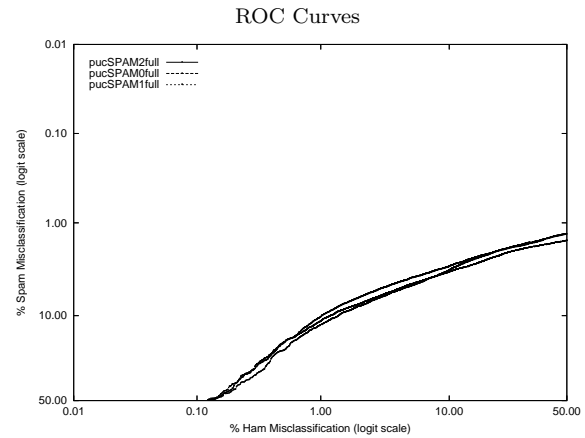


Learning Curves

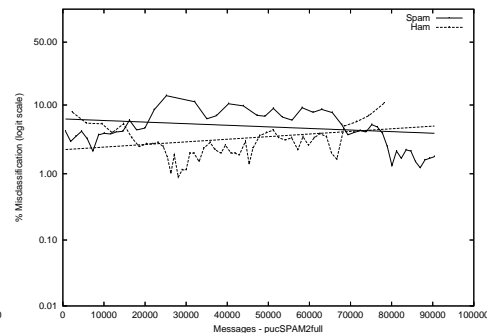
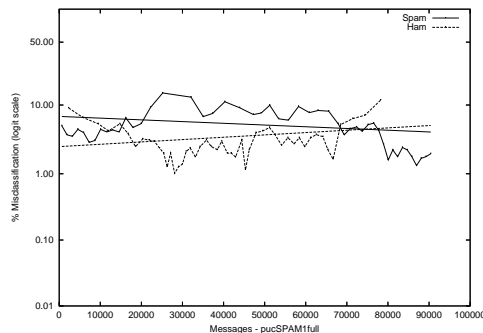
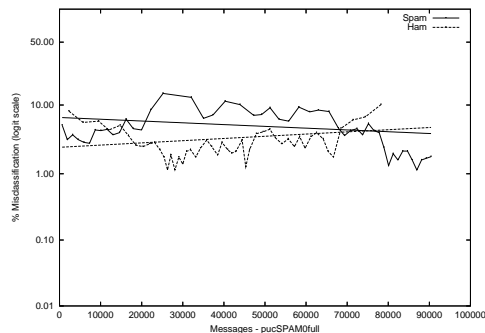


puc - PUCRS on Public trec05p-1/full Corpus

Gold Std. Filter Classification	pucSPAM0full			pucSPAM1full			pucSPAM2full		
	ham spam	ham 38054 2694	spam 1345 50096	ham spam	ham 37994 2813	spam 1405 49977	ham spam	ham 38080 2642	spam 1319 50148
Ham Misc%	3.41 (3.24-3.60)			3.57 (3.39-3.75)			3.35 (3.17-3.53)		
Spam Misc%	5.10 (4.92-5.29)			5.33 (5.14-5.52)			5.00 (4.82-5.19)		
Logit av Misc%	4.18 (4.05-4.31)			4.36 (4.23-4.50)			4.10 (3.97-4.23)		
(1-ROCA)%	2.083 (1.979 - 2.191)			2.185 (2.083 - 2.293)			1.967 (1.857 - 2.083)		
Ham Misc%	Spam Misc%								
0.01	77.61 (72.67 - 81.87)			72.36 (69.54 - 75.02)			72.26 (69.31 - 75.02)		
0.02	77.05 (74.14 - 79.72)			71.43 (70.57 - 72.27)			71.69 (68.93 - 74.30)		
0.05	64.12 (55.33 - 72.06)			61.99 (51.37 - 71.58)			63.34 (52.79 - 72.74)		
0.10	59.71 (53.22 - 65.88)			52.58 (49.59 - 55.54)			51.28 (48.23 - 54.33)		
0.20	43.06 (38.57 - 47.66)			39.52 (33.43 - 45.95)			40.82 (33.98 - 48.04)		
0.50	21.64 (19.85 - 23.55)			18.70 (16.33 - 21.34)			18.36 (16.36 - 20.55)		
1.00	12.44 (11.66 - 13.28)			11.49 (10.58 - 12.47)			10.35 (9.46 - 11.32)		
2.00	8.09 (7.76 - 8.44)			7.73 (7.37 - 8.12)			6.60 (6.27 - 6.94)		
5.00	5.02 (4.80 - 5.24)			4.80 (4.60 - 5.02)			4.13 (3.93 - 4.34)		
10.00	3.34 (3.18 - 3.52)			3.48 (3.32 - 3.65)			3.02 (2.86 - 3.19)		
20.00	2.16 (2.04 - 2.28)			2.44 (2.32 - 2.57)			2.15 (2.01 - 2.30)		
50.00	1.34 (1.23 - 1.45)			1.58 (1.48 - 1.68)			1.33 (1.22 - 1.44)		
Spam Misc%	Ham Misc%								
0.01	99.99 (nan - nan)			99.99 (nan - nan)			99.99 (nan - nan)		
0.02	99.99 (99.98 - 100.00)			99.99 (99.98 - 100.00)			99.99 (99.98 - 100.00)		
0.05	99.97 (99.94 - 99.98)			99.97 (99.94 - 99.98)			99.97 (99.94 - 99.98)		
0.10	99.94 (99.90 - 99.97)			99.94 (99.90 - 99.97)			99.94 (99.90 - 99.97)		
0.20	99.88 (99.83 - 99.92)			99.88 (99.83 - 99.92)			99.88 (99.83 - 99.92)		
0.50	99.73 (99.68 - 99.78)			99.73 (99.68 - 99.78)			99.73 (99.68 - 99.78)		
1.00	86.76 (77.18 - 92.70)			89.14 (64.92 - 97.33)			86.57 (54.85 - 97.16)		
2.00	22.60 (20.14 - 25.28)			28.95 (25.19 - 33.03)			22.86 (19.50 - 26.59)		
5.00	5.02 (4.55 - 5.53)			4.63 (4.26 - 5.03)			3.38 (3.07 - 3.71)		
10.00	1.40 (1.29 - 1.53)			1.22 (1.10 - 1.35)			1.02 (0.93 - 1.13)		
20.00	0.55 (0.48 - 0.62)			0.46 (0.39 - 0.54)			0.45 (0.39 - 0.53)		
50.00	0.14 (0.11 - 0.17)			0.12 (0.09 - 0.16)			0.12 (0.09 - 0.16)		

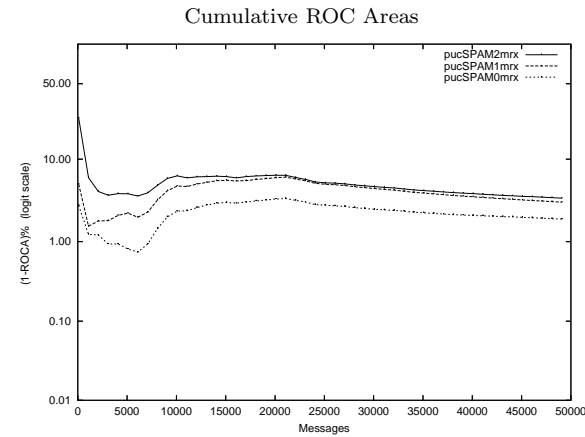
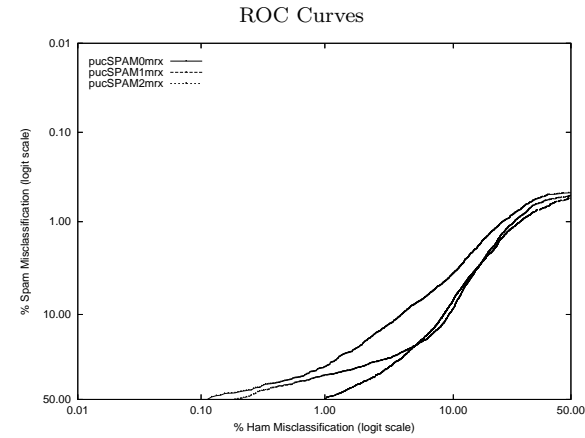


Learning Curves

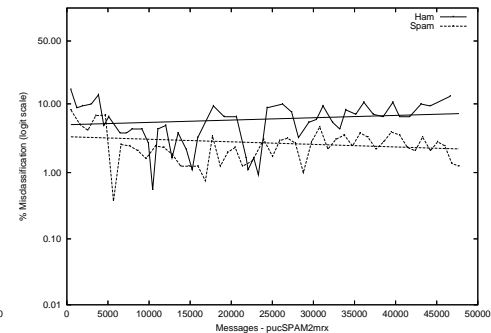
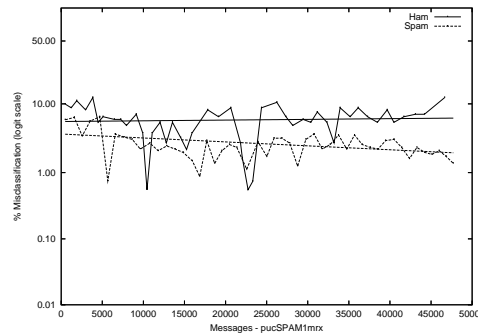
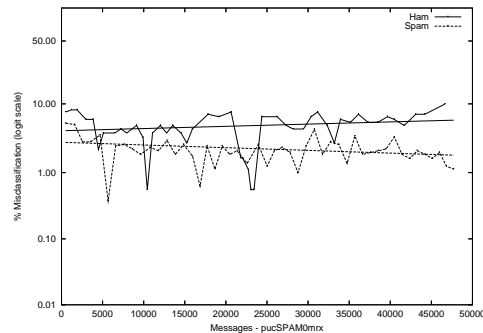


puc - PUCRS on Private Mr X. Corpus

Gold Std. Filter Classification	pucSPAM0mrx		pucSPAM1mrx		pucSPAM2mrx	
	ham	spam	ham	spam	ham	spam
Ham Misc%	8592	907	8498	1091	8489	1109
Spam Misc%	446	39141	540	38957	549	38939
Logit av Misc%	4.93 (4.50-5.40)		5.97 (5.49-6.48)		6.07 (5.59-6.59)	
(1-ROCA)%	2.26 (2.12-2.42)		2.72 (2.57-2.89)		2.77 (2.61-2.93)	
	3.35 (3.17-3.54)		4.05 (3.85-4.26)		4.12 (3.91-4.33)	
Ham Misc%	1.910 (1.783 - 2.045)		3.081 (2.904 - 3.269)		3.454 (3.290 - 3.627)	
0.01	88.84 (70.92 - 96.30)	79.05 (70.23 - 85.79)	89.24 (86.46 - 91.50)			
0.02	75.61 (50.16 - 90.52)	71.23 (56.96 - 82.24)	87.42 (83.48 - 90.53)			
0.05	58.92 (48.78 - 68.36)	59.77 (49.69 - 69.08)	82.33 (74.95 - 87.89)			
0.10	51.00 (44.28 - 57.69)	55.92 (50.34 - 61.34)	78.25 (72.60 - 83.00)			
0.20	45.28 (42.05 - 48.55)	49.62 (44.30 - 54.96)	70.43 (64.66 - 75.61)			
0.50	37.83 (35.68 - 40.03)	40.86 (38.60 - 43.16)	58.44 (53.92 - 62.82)			
1.00	30.19 (27.07 - 33.51)	35.06 (33.47 - 36.69)	49.87 (47.24 - 52.50)			
2.00	19.40 (16.74 - 22.37)	30.10 (28.48 - 31.77)	39.55 (37.40 - 41.74)			
5.00	8.05 (7.40 - 8.75)	20.41 (19.18 - 21.68)	20.89 (19.29 - 22.58)			
10.00	3.72 (3.38 - 4.10)	8.81 (7.55 - 10.26)	7.18 (6.19 - 8.32)			
20.00	1.17 (1.05 - 1.30)	1.69 (1.47 - 1.95)	1.88 (1.64 - 2.16)			
50.00	0.48 (0.42 - 0.54)	0.51 (0.45 - 0.58)	0.55 (0.49 - 0.62)			
Spam Misc%						
0.01	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)			
0.02	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)			
0.05	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)			
0.10	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)			
0.20	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)			
0.50	39.37 (nan - nan)	54.82 (nan - nan)	60.27 (nan - nan)			
1.00	21.97 (20.11 - 23.95)	25.90 (24.23 - 27.64)	28.09 (26.24 - 30.03)			
2.00	14.61 (13.69 - 15.58)	18.42 (17.40 - 19.49)	19.42 (18.38 - 20.50)			
5.00	7.99 (7.35 - 8.68)	12.31 (11.59 - 13.08)	11.85 (11.19 - 12.54)			
10.00	4.09 (3.72 - 4.50)	9.32 (8.72 - 9.95)	8.57 (8.06 - 9.12)			
20.00	1.93 (1.63 - 2.27)	5.20 (4.74 - 5.71)	5.24 (4.85 - 5.66)			
50.00	0.12 (0.06 - 0.27)	0.19 (0.11 - 0.32)	1.00 (0.81 - 1.22)			

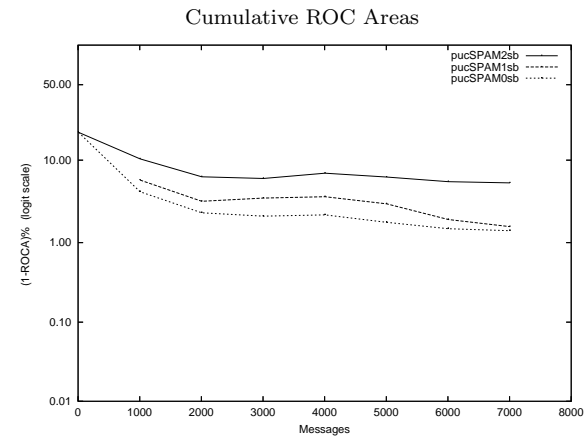
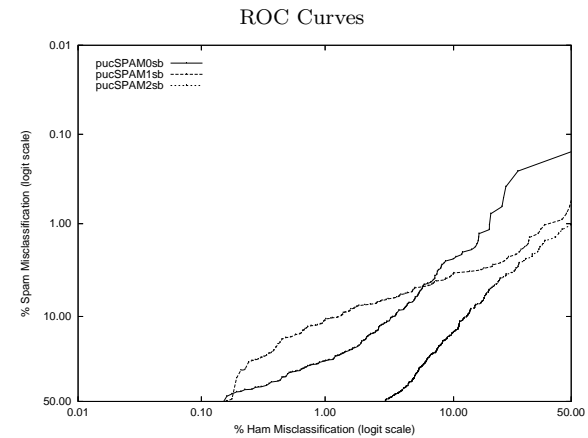


Learning Curves

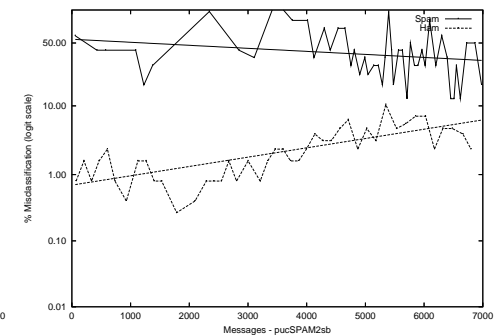
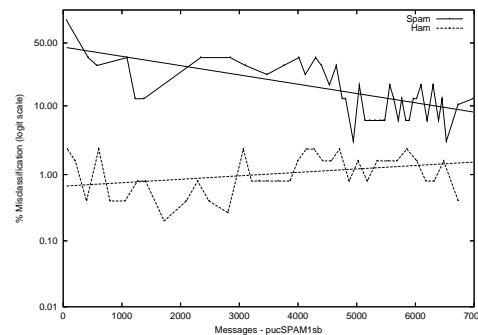
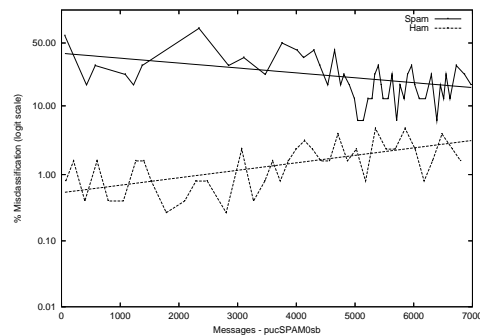


puc - PUCRS on Private S.B. Corpus

Gold Std. Filter Classification	pucSPAM0sb		pucSPAM1sb		pucSPAM2sb	
	ham	spam	ham	spam	ham	spam
Ham Misc%	6141	187	6167	134	6077	317
Spam Misc%	90	588	64	641	154	458
Logit av Misc%	1.44 (1.16-1.77)		1.03 (0.79-1.31)		2.47 (2.10-2.89)	
(1-ROCA)%	24.13 (21.16-27.30)		17.29 (14.69-20.14)		40.90 (37.42-44.46)	
	6.39 (5.64-7.23)		4.45 (3.84-5.15)		11.70 (10.63-12.85)	
	1.408 (1.116 - 1.774)		1.585 (1.082 - 2.317)		5.437 (4.767 - 6.195)	
Ham Misc%	Spam Misc%					
0.01	88.90 (75.89 - 95.33)		85.81 (72.41 - 93.30)		81.68 (76.95 - 85.62)	
0.02	76.65 (53.75 - 90.26)		76.90 (57.10 - 89.28)		79.74 (73.71 - 84.67)	
0.05	68.26 (53.07 - 80.35)		66.71 (47.52 - 81.60)		77.94 (71.51 - 83.25)	
0.10	61.81 (47.45 - 74.36)		56.52 (41.72 - 70.24)		73.42 (66.36 - 79.46)	
0.20	43.87 (33.47 - 54.85)		35.35 (17.34 - 58.78)		70.19 (65.20 - 74.75)	
0.50	33.03 (27.35 - 39.25)		16.13 (11.62 - 21.95)		66.06 (62.35 - 69.59)	
1.00	26.06 (22.12 - 30.44)		10.97 (8.44 - 14.14)		61.81 (57.29 - 66.13)	
2.00	17.42 (13.55 - 22.11)		7.74 (5.87 - 10.15)		57.55 (53.15 - 61.83)	
5.00	7.10 (5.14 - 9.73)		5.29 (3.76 - 7.39)		35.74 (29.70 - 42.28)	
10.00	2.45 (1.49 - 4.01)		3.61 (2.28 - 5.67)		14.32 (11.36 - 17.90)	
20.00	0.77 (0.30 - 1.99)		2.84 (1.71 - 4.69)		4.65 (3.27 - 6.56)	
50.00	0.26 (0.07 - 0.96)		0.65 (0.19 - 2.15)		1.03 (0.47 - 2.24)	
Spam Misc%	Ham Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.05	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.10	66.46 (42.11 - 84.37)		62.67 (55.41 - 69.40)		85.76 (76.16 - 91.91)	
0.20	59.54 (23.37 - 87.65)		56.89 (47.87 - 65.48)		78.67 (60.16 - 90.01)	
0.50	22.79 (8.06 - 49.83)		51.65 (41.24 - 61.91)		75.48 (53.84 - 89.04)	
1.00	18.22 (13.33 - 24.39)		45.66 (30.89 - 61.23)		52.27 (36.52 - 67.58)	
2.00	13.45 (8.33 - 20.99)		28.17 (16.18 - 44.33)		35.32 (25.25 - 46.89)	
5.00	5.87 (4.61 - 7.46)		5.86 (2.69 - 12.29)		18.89 (15.35 - 23.03)	
10.00	3.59 (2.81 - 4.59)		1.35 (0.81 - 2.25)		12.52 (10.70 - 14.60)	
20.00	1.72 (1.26 - 2.34)		0.40 (0.27 - 0.60)		7.74 (6.62 - 9.02)	
50.00	0.16 (0.08 - 0.30)		0.16 (0.08 - 0.34)		3.02 (2.45 - 3.70)	

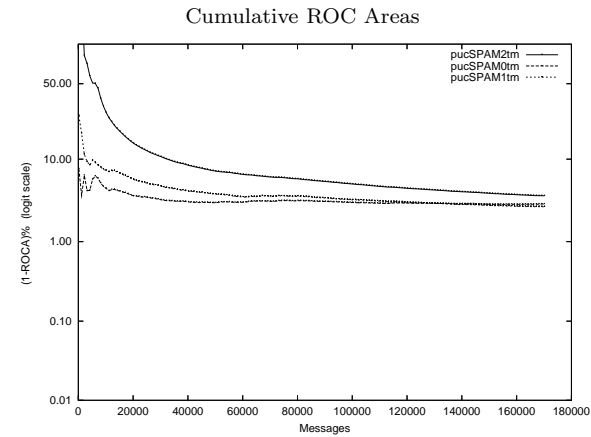
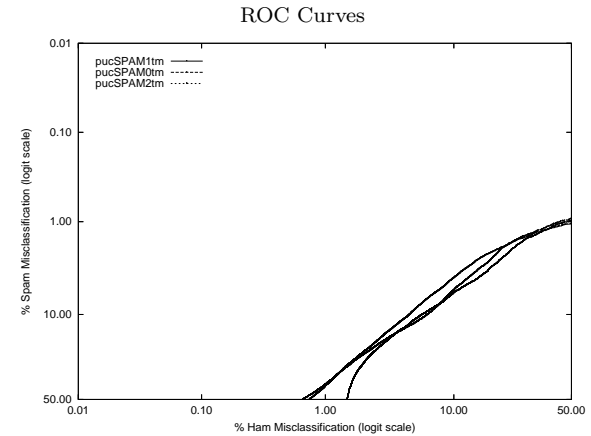


Learning Curves

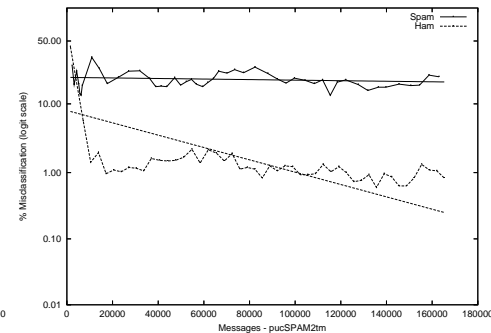
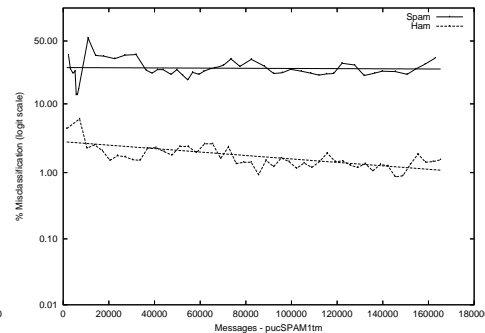
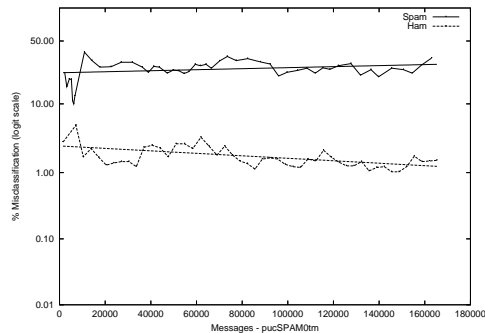


puc - PUCRS on Private T.M. Corpus

Gold Std. Filter Classification	pucSPAM0tm		pucSPAM1tm		pucSPAM2tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	1.77 (1.70-1.84)		1.80 (1.74-1.87)		2.17 (2.10-2.25)	
Spam Misc%	27.34 (26.71-27.97)		27.87 (27.24-28.50)		20.73 (20.16-21.31)	
Logit av Misc%	7.61 (7.43-7.78)		7.77 (7.60-7.95)		7.08 (6.92-7.24)	
(1-ROCA)%	2.925 (2.779 - 3.078)		2.712 (2.588 - 2.842)		3.688 (3.543 - 3.838)	
Ham Misc%	Spam Misc%					
0.01	98.13 (96.96 - 98.86)		97.14 (96.21 - 97.85)		100.00 (nan - nan)	
0.02	96.81 (95.70 - 97.64)		95.95 (94.85 - 96.82)		99.99 (nan - nan)	
0.05	93.56 (92.15 - 94.73)		93.08 (91.92 - 94.08)		99.99 (nan - nan)	
0.10	88.94 (87.49 - 90.25)		88.48 (86.90 - 89.88)		99.99 (nan - nan)	
0.20	77.76 (75.89 - 79.52)		78.96 (76.60 - 81.14)		99.99 (nan - nan)	
0.50	56.57 (54.65 - 58.48)		59.59 (58.05 - 61.11)		99.96 (99.92 - 99.98)	
1.00	40.29 (38.83 - 41.77)		41.66 (39.79 - 43.56)		93.42 (91.53 - 94.91)	
2.00	24.10 (23.28 - 24.94)		23.04 (22.21 - 23.89)		28.61 (27.10 - 30.17)	
5.00	12.41 (11.90 - 12.94)		9.10 (8.68 - 9.54)		11.50 (11.05 - 11.96)	
10.00	5.60 (5.27 - 5.94)		4.14 (3.89 - 4.40)		6.00 (5.67 - 6.34)	
20.00	2.24 (2.05 - 2.46)		2.03 (1.85 - 2.24)		2.85 (2.63 - 3.08)	
50.00	0.93 (0.80 - 1.08)		1.05 (0.92 - 1.20)		0.98 (0.85 - 1.13)	
Spam Misc%	Ham Misc%					
0.01	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.02	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.05	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.10	100.00 (nan - nan)		100.00 (nan - nan)		100.00 (nan - nan)	
0.20	99.99 (99.98 - 100.00)		99.99 (99.98 - 100.00)		99.99 (99.98 - 100.00)	
0.50	91.73 (73.08 - 97.84)		95.26 (80.07 - 99.01)		96.42 (81.83 - 99.38)	
1.00	44.87 (36.62 - 53.41)		54.60 (44.29 - 64.53)		48.85 (41.16 - 56.60)	
2.00	21.20 (19.68 - 22.82)		20.66 (18.21 - 23.34)		24.95 (23.40 - 26.56)	
5.00	10.89 (10.30 - 11.51)		8.62 (8.25 - 9.01)		12.21 (11.39 - 13.09)	
10.00	6.38 (6.11 - 6.65)		4.64 (4.46 - 4.83)		5.83 (5.51 - 6.16)	
20.00	2.55 (2.44 - 2.67)		2.33 (2.24 - 2.43)		2.78 (2.66 - 2.90)	
50.00	0.66 (0.61 - 0.71)		0.74 (0.69 - 0.79)		1.49 (1.43 - 1.56)	



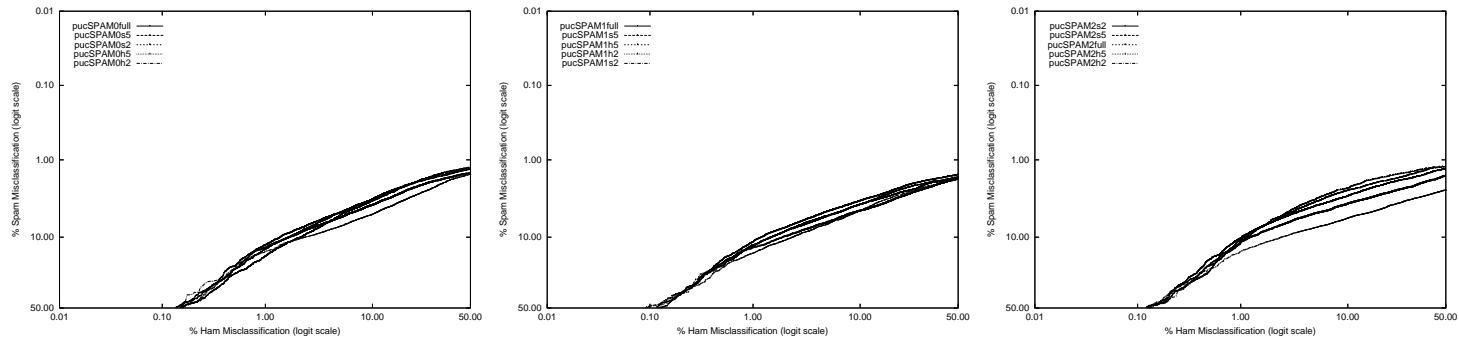
Learning Curves



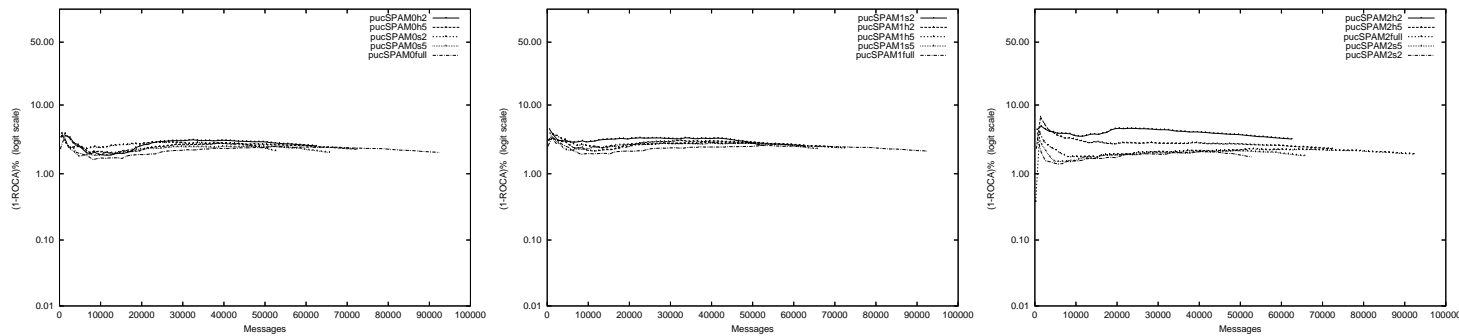
puc - PUCRS - Public subset Corpora

	pucSPAM0full	pucSPAM0h2	pucSPAM0h5	pucSPAM0s2	pucSPAM0s5
Ham Misc%	3.41 (3.24-3.60)	9.62 (9.04-10.22)	5.82 (5.49-6.15)	1.62 (1.50-1.75)	2.28 (2.14-2.44)
Spam Misc%	5.10 (4.92-5.29)	3.57 (3.42-3.73)	4.32 (4.15-4.50)	9.70 (9.20-10.22)	6.86 (6.56-7.17)
Logit av Misc% (1-ROCA)%	2.083 (1.979 - 2.191)	2.600 (2.471 - 2.734)	2.314 (2.204 - 2.429)	2.200 (2.034 - 2.378)	2.083 (1.953 - 2.222)
	pucSPAM1full	pucSPAM1h2	pucSPAM1h5	pucSPAM1s2	pucSPAM1s5
Ham Misc%	3.57 (3.39-3.75)	10.07 (9.48-10.69)	6.06 (5.73-6.40)	1.71 (1.58-1.84)	2.44 (2.29-2.59)
Spam Misc%	5.33 (5.14-5.52)	3.74 (3.58-3.90)	4.50 (4.32-4.68)	10.25 (9.74-10.78)	7.31 (6.99-7.63)
Logit av Misc% (1-ROCA)%	2.185 (2.083 - 2.293)	2.618 (2.490 - 2.753)	2.409 (2.295 - 2.528)	2.623 (2.424 - 2.838)	2.367 (2.230 - 2.513)
	pucSPAM2full	pucSPAM2h2	pucSPAM2h5	pucSPAM2s2	pucSPAM2s5
Ham Misc%	3.35 (3.17-3.53)	10.51 (9.91-11.14)	6.00 (5.67-6.34)	1.50 (1.38-1.62)	2.15 (2.01-2.30)
Spam Misc%	5.00 (4.82-5.19)	3.92 (3.76-4.09)	4.46 (4.29-4.64)	8.97 (8.49-9.47)	6.47 (6.18-6.78)
Logit av Misc% (1-ROCA)%	1.967 (1.857 - 2.083)	3.274 (3.147 - 3.407)	2.358 (2.242 - 2.480)	1.788 (1.611 - 1.984)	1.853 (1.729 - 1.986)

ROC Curves

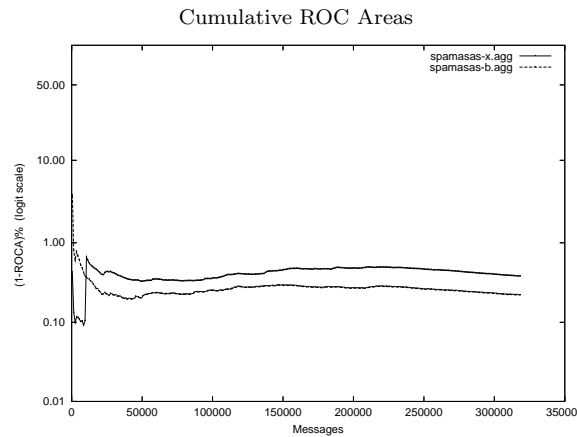
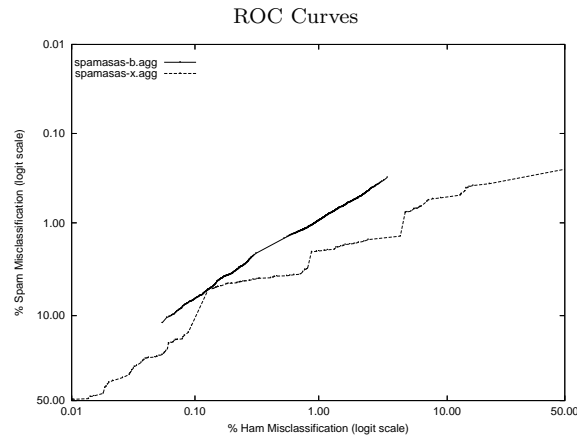


Cumulative ROC Area

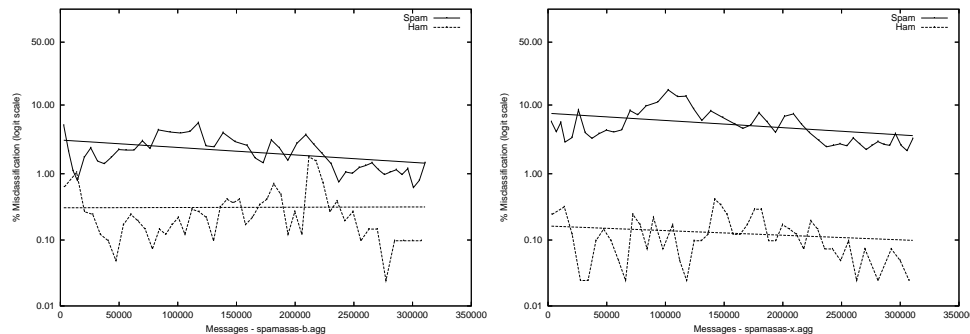


spamasas - Spmassassin (non-participant) on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	spamasas-b.agg		spamasas-x.agg	
	ham	spam	ham	spam
ham	204714	2454	205091	6099
spam	640	110675	262	107030
Ham Misc%	0.31 (0.29-0.34)		0.13 (0.11-0.14)	
Spam Misc%	2.17 (2.09-2.26)		5.39 (5.26-5.52)	
Logit av Misc%	0.83 (0.79-0.86)		0.85 (0.80-0.90)	
(1-ROCA)%	0.220 (0.202 - 0.240)		0.380 (0.352 - 0.411)	
Ham Misc%	Spam Misc%			
0.01	81.23 (72.50 - 87.66)	50.08 (48.11 - 52.05)	38.38 (30.63 - 46.76)	23.94 (21.77 - 26.25)
0.02	68.89 (58.18 - 77.89)	38.38 (30.63 - 46.76)	23.94 (21.77 - 26.25)	11.17 (8.86 - 13.98)
0.05	19.73 (9.65 - 36.15)	11.17 (8.86 - 13.98)	4.58 (4.46 - 4.71)	3.84 (3.73 - 3.95)
0.10	6.72 (5.99 - 7.54)	4.58 (4.46 - 4.71)	2.04 (1.96 - 2.13)	1.67 (1.60 - 1.74)
0.20	3.59 (3.28 - 3.93)	4.58 (4.46 - 4.71)	1.67 (1.60 - 1.74)	0.75 (0.71 - 0.80)
0.50	1.57 (1.45 - 1.70)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	0.52 (0.48 - 0.56)
1.00	0.94 (0.87 - 1.01)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	0.36 (0.33 - 0.40)
2.00	0.54 (0.50 - 0.58)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	0.27 (0.24 - 0.31)
5.00	0.31 (0.28 - 0.34)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	
10.00	0.29 (0.26 - 0.33)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	
20.00	0.26 (0.23 - 0.29)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	
50.00	0.17 (0.15 - 0.20)	3.84 (3.73 - 3.95)	1.67 (1.60 - 1.74)	
Spam Misc%	Ham Misc%			
0.01	96.05 (93.30 - 97.69)	99.94 (99.93 - 99.95)	99.94 (99.92 - 99.95)	99.91 (99.90 - 99.92)
0.02	93.47 (90.96 - 95.32)	99.94 (99.92 - 99.95)	99.91 (99.90 - 99.92)	73.86 (64.19 - 81.67)
0.05	85.76 (81.61 - 89.09)	99.91 (99.90 - 99.92)	73.86 (64.19 - 81.67)	11.78 (9.14 - 15.07)
0.10	70.92 (65.71 - 75.62)	99.91 (95.69 - 100.00)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)
0.20	39.09 (30.59 - 48.31)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	1.14 (0.92 - 1.42)
0.50	2.19 (1.98 - 2.42)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	0.15 (0.14 - 0.17)
1.00	0.94 (0.86 - 1.02)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	0.11 (0.10 - 0.12)
2.00	0.37 (0.33 - 0.41)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	0.06 (0.05 - 0.07)
5.00	0.14 (0.12 - 0.16)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	0.06 (0.05 - 0.07)
10.00	0.06 (0.05 - 0.08)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	0.03 (0.02 - 0.04)
20.00	0.05 (0.04 - 0.06)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	
50.00	0.03 (0.02 - 0.04)	11.78 (9.14 - 15.07)	4.67 (4.57 - 4.77)	

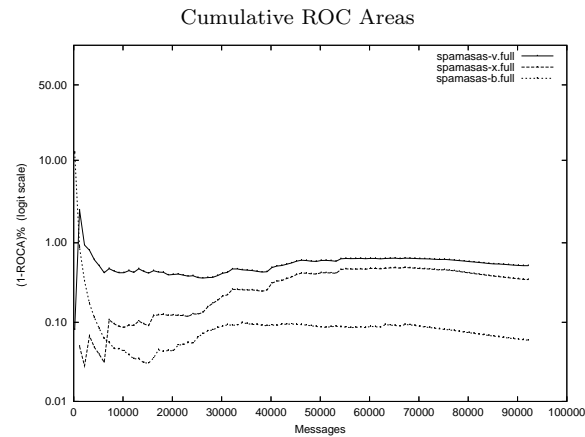
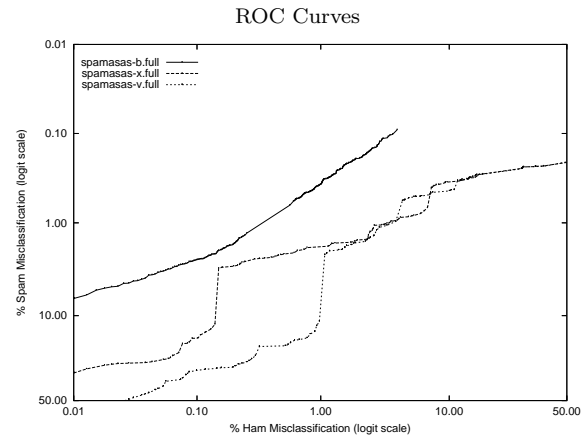


Learning Curves

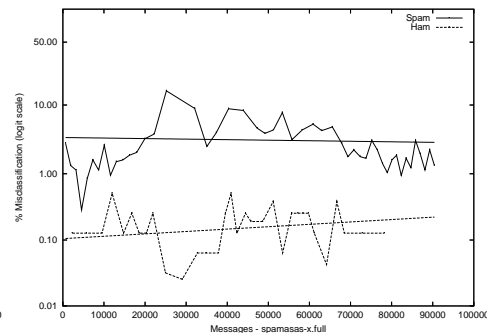
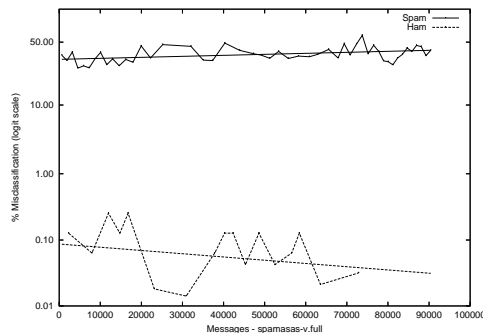
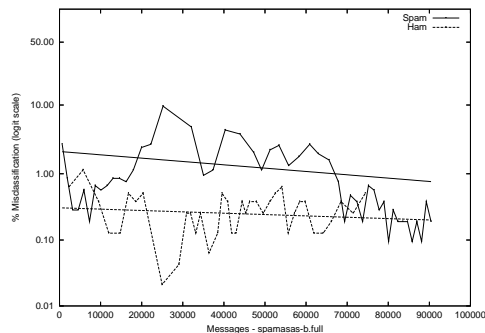


spamasas - Spmassassin (non-participant) on Public trec05p-1/full Corpus

Gold Std.	spamasas-b.full			spamasas-v.full			spamasas-x.full			
	Filter	ham	spam	ham	spam	ham	spam	ham	spam	
Classification	ham spam	39299 100	683 52107	ham spam	39377 22	20857 31933	ham spam	39340 59	1670 51120	
Ham Misc%		0.25 (0.21-0.31)			0.06 (0.03-0.08)			0.15 (0.11-0.19)		
Spam Misc%		1.29 (1.20-1.39)			39.51 (39.09-39.93)			3.16 (3.02-3.32)		
Logit av Misc%		0.57 (0.52-0.64)			1.87 (1.53-2.30)			0.70 (0.61-0.79)		
(1-ROCA)%		0.059 (0.044 - 0.081)			0.516 (0.464 - 0.574)			0.345 (0.305 - 0.390)		
Ham Misc%		Spam Misc%								
0.01		7.81 (2.25 - 23.74)			56.62 (51.79 - 61.34)			38.21 (27.07 - 50.74)		
0.02		5.19 (3.75 - 7.14)			54.84 (49.73 - 59.85)			28.41 (24.57 - 32.58)		
0.05		3.51 (2.88 - 4.26)			42.10 (34.85 - 49.70)			26.54 (22.11 - 31.51)		
0.10		2.56 (2.26 - 2.90)			31.31 (28.66 - 34.10)			16.59 (10.50 - 25.21)		
0.20		1.71 (1.38 - 2.11)			29.42 (27.32 - 31.62)			3.00 (2.17 - 4.13)		
0.50		0.78 (0.64 - 0.95)			19.55 (18.65 - 20.49)			2.22 (2.07 - 2.37)		
1.00		0.37 (0.30 - 0.46)			9.33 (3.37 - 23.30)			1.83 (1.72 - 1.95)		
2.00		0.19 (0.15 - 0.24)			1.64 (1.53 - 1.76)			1.55 (1.44 - 1.67)		
5.00		0.09 (0.06 - 0.12)			0.53 (0.46 - 0.60)			0.85 (0.77 - 0.94)		
10.00		0.09 (0.06 - 0.12)			0.44 (0.38 - 0.50)			0.35 (0.31 - 0.40)		
20.00		0.08 (0.05 - 0.11)			0.27 (0.23 - 0.33)			0.27 (0.23 - 0.32)		
50.00		0.04 (0.02 - 0.06)			0.23 (0.19 - 0.28)			0.23 (0.19 - 0.28)		
Spam Misc%		Ham Misc%								
0.01		78.00 (57.08 - 90.43)			99.66 (99.60 - 99.71)			99.79 (99.74 - 99.83)		
0.02		72.90 (56.40 - 84.84)			99.65 (99.59 - 99.70)			99.78 (99.72 - 99.82)		
0.05		42.98 (22.35 - 66.37)			99.64 (99.57 - 99.69)			99.76 (99.65 - 99.84)		
0.10		3.77 (0.96 - 13.64)			99.64 (99.57 - 99.69)			99.64 (99.57 - 99.70)		
0.20		1.85 (1.34 - 2.55)			99.64 (91.66 - 99.99)			99.64 (90.17 - 99.99)		
0.50		0.71 (0.60 - 0.84)			6.30 (4.21 - 9.34)			7.22 (6.95 - 7.49)		
1.00		0.40 (0.34 - 0.47)			3.80 (3.35 - 4.31)			3.51 (2.84 - 4.32)		
2.00		0.15 (0.12 - 0.20)			1.32 (1.04 - 1.69)			0.68 (0.56 - 0.83)		
5.00		0.02 (0.01 - 0.06)			1.05 (0.95 - 1.15)			0.15 (0.11 - 0.20)		
10.00		0.01 (0.00 - 0.02)			0.99 (0.90 - 1.09)			0.14 (0.11 - 0.19)		
20.00		0.00 (0.00 - 0.01)			0.32 (0.25 - 0.40)			0.07 (0.05 - 0.11)		
50.00		0.00 (0.00 - 0.01)			0.03 (0.01 - 0.04)			0.00 (0.00 - 0.01)		

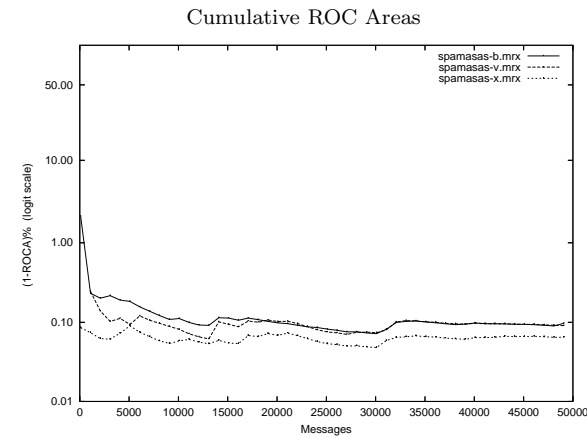
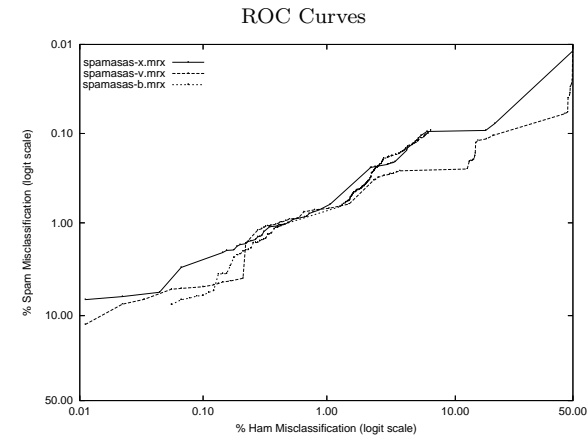


Learning Curves

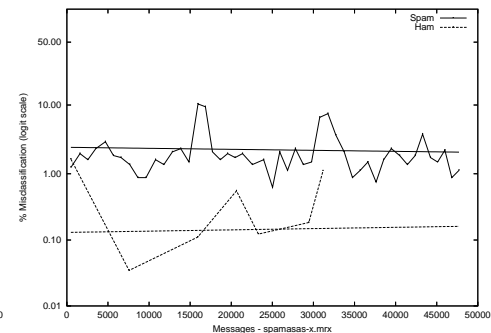
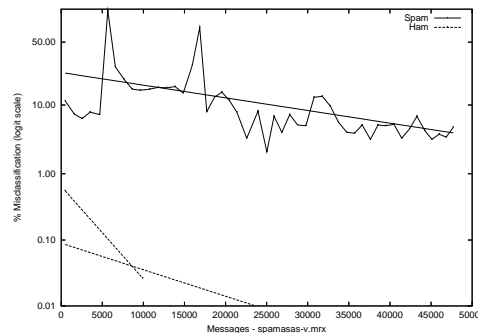
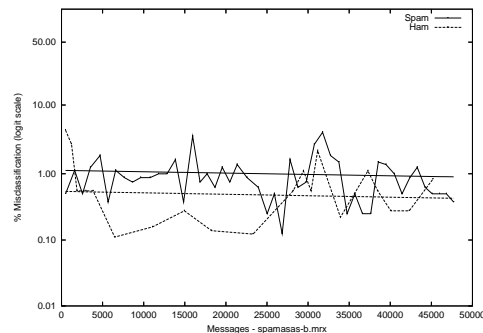


spamasas - Spmassassin (non-participant) on Private Mr X. Corpus

Gold Std. Filter Classification	spamasas-b.mrx		spamasas-v.mrx		spamasas-x.mrx	
	ham	spam	ham	spam	ham	spam
Classification	8994	401	9036	4686	9025	912
	spam	44	39647	spam	2	35362
Ham Misc%	0.49 (0.35-0.65)		0.02 (0.00-0.08)		0.14 (0.08-0.25)	
Spam Misc%	1.00 (0.91-1.10)		11.70 (11.39-12.02)		2.28 (2.13-2.43)	
Logit av Misc%	0.70 (0.60-0.82)		0.54 (0.27-1.07)		0.58 (0.44-0.76)	
(1-ROCA)%	0.097 (0.070 - 0.135)		0.091 (0.075 - 0.110)		0.065 (0.049 - 0.086)	
Ham Misc%	Spam Misc%					
0.01	83.06 (20.42 - 98.94)		27.22 (9.25 - 57.84)		16.09 (6.49 - 34.62)	
0.02	74.69 (16.77 - 97.74)		12.24 (4.10 - 31.25)		6.84 (2.36 - 18.24)	
0.05	12.82 (2.69 - 43.85)		6.66 (3.54 - 12.19)		3.56 (1.84 - 6.77)	
0.10	6.19 (3.27 - 11.39)		4.97 (3.82 - 6.46)		2.50 (1.81 - 3.44)	
0.20	2.21 (1.26 - 3.87)		4.11 (1.43 - 11.26)		1.73 (1.26 - 2.37)	
0.50	0.99 (0.84 - 1.17)		0.91 (0.78 - 1.06)		0.98 (0.82 - 1.17)	
1.00	0.77 (0.65 - 0.91)		0.69 (0.61 - 0.79)		0.65 (0.53 - 0.80)	
2.00	0.38 (0.29 - 0.49)		0.46 (0.36 - 0.60)		0.33 (0.23 - 0.47)	
5.00	0.13 (0.10 - 0.17)		0.26 (0.21 - 0.32)		0.12 (0.09 - 0.17)	
10.00	0.09 (0.06 - 0.13)		0.25 (0.21 - 0.30)		0.09 (0.07 - 0.14)	
20.00	0.08 (0.05 - 0.12)		0.10 (0.07 - 0.14)		0.07 (0.05 - 0.11)	
50.00	0.05 (0.03 - 0.09)		0.01 (0.00 - 0.05)		0.04 (0.03 - 0.08)	
Spam Misc%	Ham Misc%					
0.01	85.18 (66.29 - 94.39)		49.83 (24.75 - 75.00)		70.07 (63.00 - 76.30)	
0.02	81.41 (65.96 - 90.82)		49.72 (44.96 - 54.49)		66.38 (55.16 - 76.01)	
0.05	55.00 (26.67 - 80.42)		47.59 (40.29 - 54.99)		41.72 (22.17 - 64.29)	
0.10	6.09 (2.10 - 16.39)		20.69 (10.92 - 35.69)		5.75 (1.53 - 19.33)	
0.20	2.79 (2.17 - 3.57)		13.21 (10.06 - 17.17)		3.55 (2.55 - 4.92)	
0.50	1.57 (1.32 - 1.87)		1.87 (1.56 - 2.24)		1.45 (1.11 - 1.90)	
1.00	0.50 (0.33 - 0.76)		0.42 (0.30 - 0.58)		0.49 (0.32 - 0.73)	
2.00	0.23 (0.15 - 0.36)		0.22 (0.14 - 0.34)		0.17 (0.10 - 0.29)	
5.00	0.13 (0.08 - 0.23)		0.10 (0.04 - 0.23)		0.04 (0.01 - 0.16)	
10.00	0.06 (0.02 - 0.19)		0.02 (0.01 - 0.10)		0.01 (0.00 - 0.04)	
20.00	0.04 (0.01 - 0.17)		0.01 (0.00 - 0.04)		0.01 (0.01 - 0.01)	
50.00	0.02 (0.01 - 0.08)		0.01 (0.01 - 0.01)		0.01 (0.01 - 0.01)	

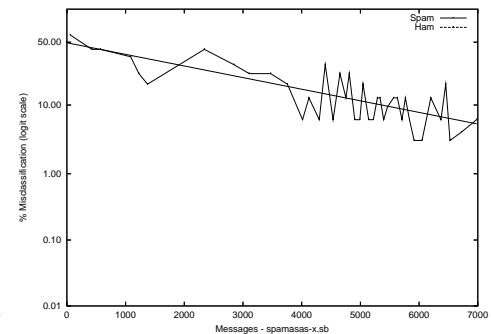
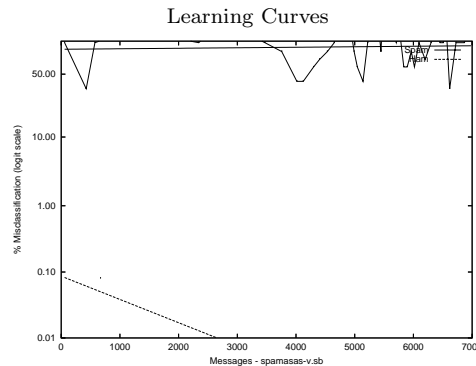
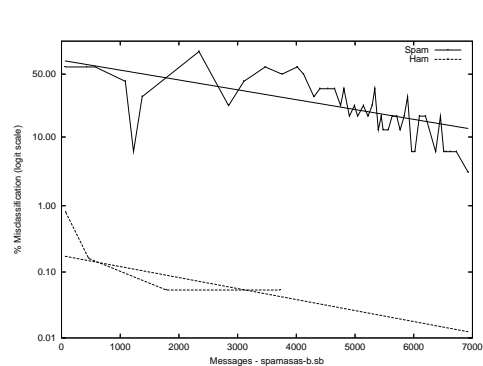
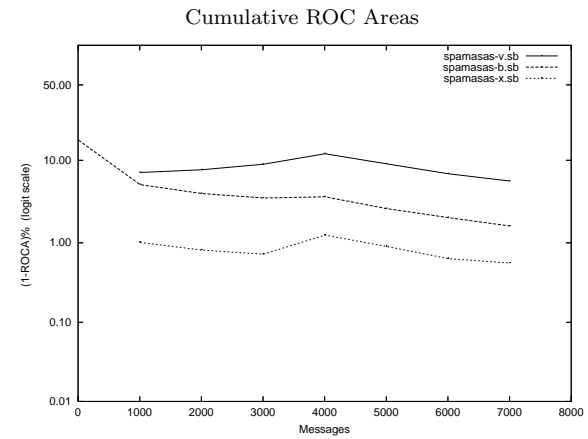
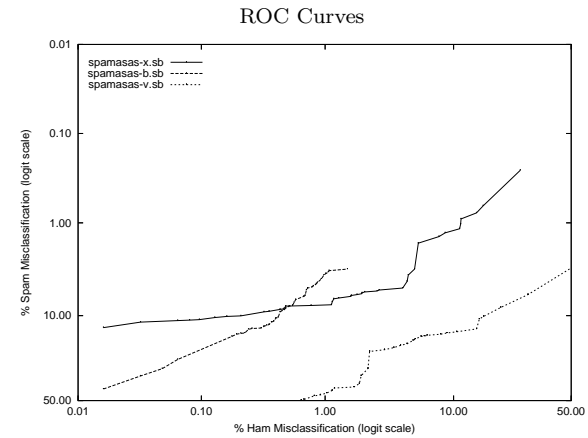


Learning Curves



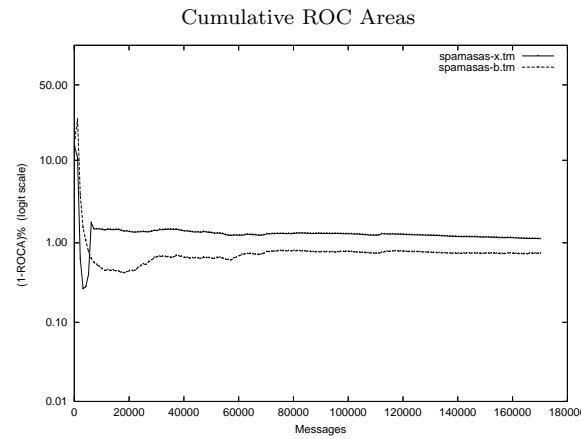
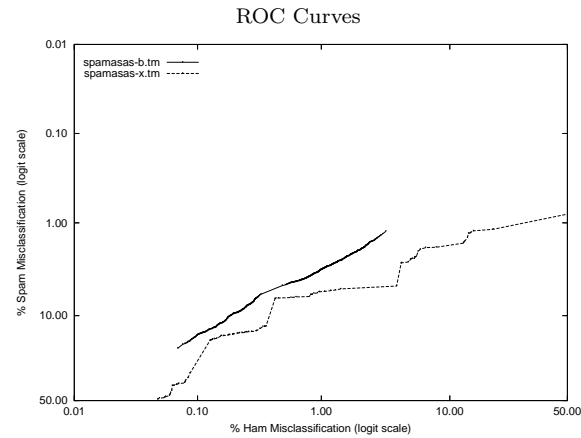
spamasas - Spamassassin (non-participant) on Private S.B. Corpus

Gold Std. Filter Classification	spamasas-b.sb		spamasas-v.sb		spamasas-x.sb	
	ham	spam	ham	spam	ham	spam
Ham Misc%	0.06	(0.02-0.16)	0.02	(0.00-0.09)	0.00	(0.00-0.05)
Spam Misc%	25.68	(22.63-28.91)	72.13	(68.83-75.26)	14.84	(12.41-17.54)
Logit av Misc%	1.47	(0.90-2.39)	2.00	(0.76-5.17)	0.00	(0.00-nan)
(1-ROCA)%	1.620	(1.054 - 2.481)	5.736	(4.840 - 6.785)	0.558	(0.359 - 0.865)
Ham Misc%	Spam Misc%					
0.01	50.06	(35.45 - 64.66)	73.81	(68.21 - 78.72)	13.68	(10.76 - 17.24)
0.02	42.58	(26.12 - 60.86)	68.52	(62.44 - 74.02)	13.03	(9.94 - 16.91)
0.05	30.45	(17.31 - 47.80)	68.26	(65.10 - 71.26)	11.23	(8.87 - 14.12)
0.10	19.87	(12.61 - 29.88)	68.26	(64.33 - 71.95)	10.84	(8.63 - 13.52)
0.20	14.97	(10.93 - 20.16)	68.26	(60.29 - 75.28)	10.06	(8.04 - 12.54)
0.50	8.39	(5.60 - 12.37)	52.65	(47.33 - 57.90)	8.00	(6.30 - 10.11)
1.00	3.74	(2.41 - 5.77)	45.16	(40.69 - 49.71)	7.74	(5.96 - 9.99)
2.00	3.23	(2.21 - 4.69)	33.55	(23.02 - 46.01)	5.81	(4.40 - 7.62)
5.00	3.23	(2.20 - 4.70)	17.42	(14.61 - 20.64)	3.35	(1.92 - 5.81)
10.00	3.23	(2.17 - 4.78)	14.58	(12.31 - 17.18)	1.29	(0.57 - 2.90)
20.00	2.84	(1.85 - 4.33)	8.65	(6.80 - 10.93)	0.65	(0.17 - 2.43)
50.00	1.29	(0.69 - 2.39)	3.74	(2.55 - 5.47)	0.13	(0.05 - 0.36)
Spam Misc%	Ham Misc%					
0.01	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.02	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.05	100.00	(nan - nan)	100.00	(nan - nan)	100.00	(nan - nan)
0.10	99.10	(86.81 - 99.95)	99.49	(73.67 - 99.99)	53.12	(17.35 - 85.95)
0.20	97.87	(81.97 - 99.78)	72.36	(5.88 - 99.10)	42.55	(16.97 - 72.85)
0.50	92.35	(80.07 - 97.31)	69.97	(65.20 - 74.35)	26.06	(12.20 - 47.21)
1.00	57.46	(34.42 - 77.66)	66.33	(64.40 - 68.21)	11.33	(5.51 - 21.87)
2.00	33.02	(12.71 - 62.54)	63.60	(56.41 - 70.23)	5.39	(3.73 - 7.73)
5.00	0.80	(0.56 - 1.16)	36.40	(25.64 - 48.72)	4.29	(1.93 - 9.24)
10.00	0.43	(0.29 - 0.65)	17.01	(13.59 - 21.09)	0.29	(0.04 - 2.28)
20.00	0.10	(0.04 - 0.22)	3.61	(2.12 - 6.08)	0.01	(0.01 - 0.01)
50.00	0.02	(0.01 - 0.05)	0.64	(0.41 - 1.01)	0.01	(0.01 - 0.01)

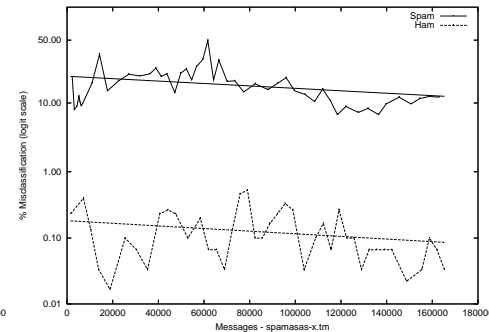
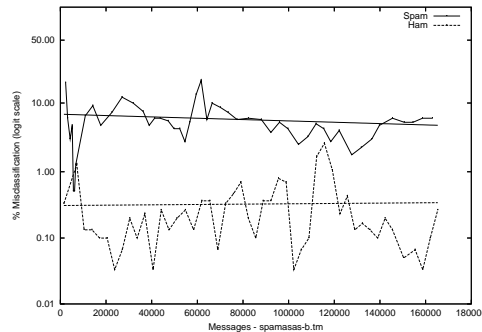


spamasas - Spmassassin (non-participant) on Private T.M. Corpus

Gold Std. Filter Classification	spamasas-b.tm		spamasas-x.tm	
	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.33 (0.30-0.36)	150193	0.13 (0.11-0.15)	150495
Spam Misc%	6.00 (5.67-6.34)	1171	17.43 (16.90-17.97)	3402
Logit av Misc%	1.43 (1.35-1.50)	492	1.61 (1.49-1.73)	18345
(1-ROCA)%	0.736 (0.648 - 0.837)		1.123 (1.019 - 1.237)	
Ham Misc%	Spam Misc%			
0.01	88.94 (83.00 - 92.97)		70.54 (69.79 - 71.27)	
0.02	76.53 (67.72 - 83.52)		67.86 (63.04 - 72.33)	
0.05	45.75 (33.32 - 58.73)		48.36 (43.81 - 52.93)	
0.10	15.58 (13.71 - 17.65)		29.50 (24.36 - 35.21)	
0.20	9.47 (8.78 - 10.20)		15.06 (14.51 - 15.62)	
0.50	4.80 (4.49 - 5.13)		6.53 (6.21 - 6.87)	
1.00	3.27 (3.02 - 3.55)		5.66 (5.36 - 5.97)	
2.00	2.02 (1.83 - 2.23)		5.20 (4.92 - 5.49)	
5.00	1.21 (1.07 - 1.37)		2.59 (2.36 - 2.85)	
10.00	1.14 (1.01 - 1.30)		1.79 (1.60 - 2.00)	
20.00	1.00 (0.87 - 1.16)		1.18 (1.04 - 1.33)	
50.00	0.65 (0.54 - 0.79)		0.84 (0.72 - 0.97)	
Spam Misc%	Ham Misc%			
0.01	99.35 (95.88 - 99.90)		98.98 (98.04 - 99.47)	
0.02	98.40 (95.71 - 99.41)		98.10 (97.45 - 98.58)	
0.05	96.48 (93.60 - 98.08)		98.09 (98.01 - 98.16)	
0.10	90.73 (85.88 - 94.03)		98.06 (97.98 - 98.14)	
0.20	81.22 (74.54 - 86.46)		98.02 (97.37 - 98.51)	
0.50	62.16 (53.74 - 69.90)		80.63 (70.02 - 88.12)	
1.00	20.13 (11.17 - 33.56)		32.34 (21.21 - 45.90)	
2.00	2.04 (1.79 - 2.33)		5.93 (4.82 - 7.29)	
5.00	0.46 (0.41 - 0.52)		3.42 (1.91 - 6.03)	
10.00	0.19 (0.16 - 0.21)		0.38 (0.35 - 0.42)	
20.00	0.07 (0.06 - 0.09)		0.12 (0.11 - 0.14)	
50.00	0.05 (0.04 - 0.06)		0.05 (0.04 - 0.06)	

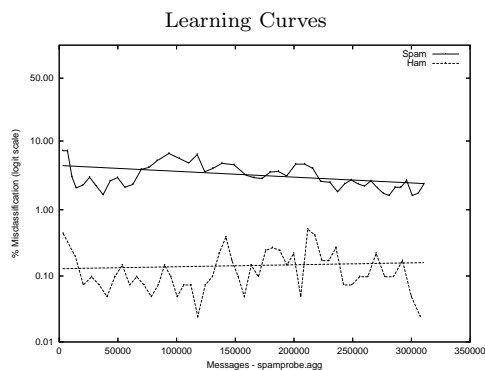
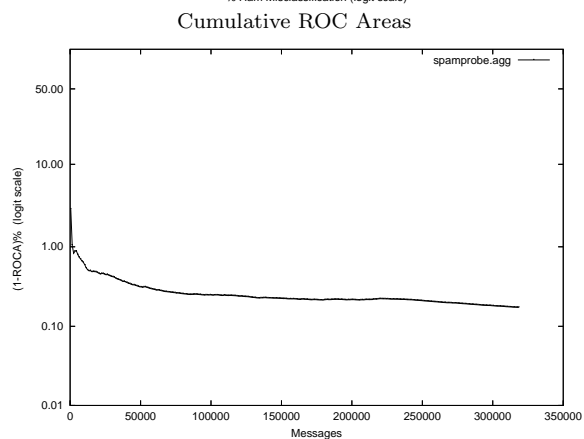
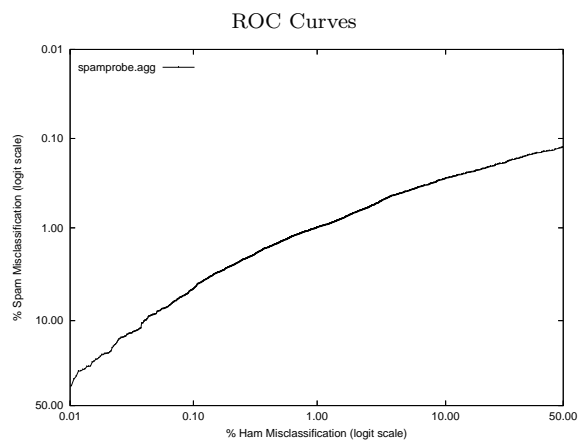


Learning Curves



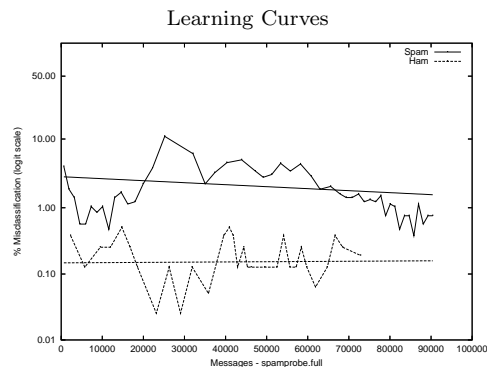
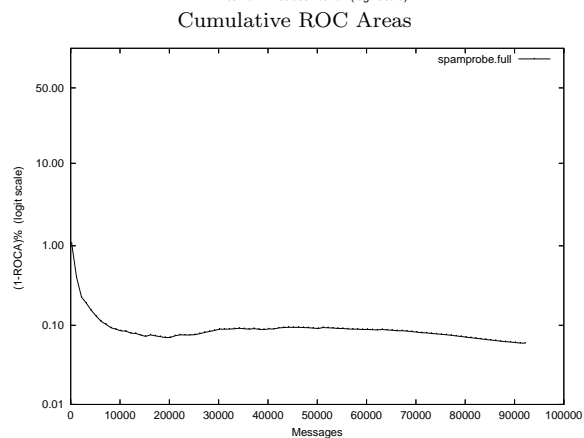
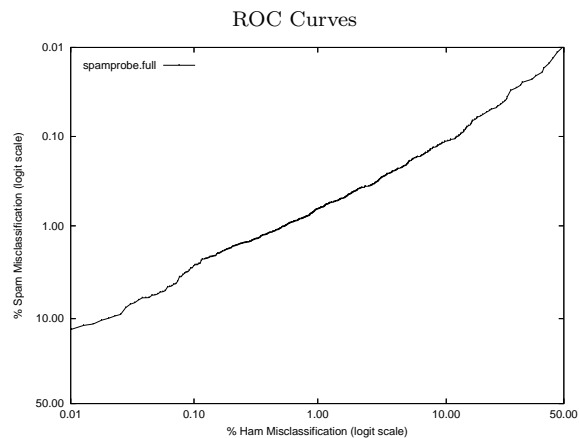
spamprobe - Spamprobe (non-partipant) on Aggregate of full, mrx, sb, tm Corpus

Gold Std.	spamprobe.agg	
	ham	spam
Filter	205058	3793
Classification	spam	295 109336
Ham Misc%	0.14 (0.13-0.16)	
Spam Misc%	3.35 (3.25-3.46)	
Logit av Misc%	0.70 (0.66-0.74)	
(1-ROCA)%	0.173 (0.158 - 0.190)	
Ham Misc%	Spam Misc%	
0.01	37.79 (20.22 - 59.28)	
0.02	20.54 (14.40 - 28.42)	
0.05	8.56 (7.28 - 10.05)	
0.10	4.71 (4.03 - 5.51)	
0.20	2.67 (2.50 - 2.86)	
0.50	1.45 (1.35 - 1.55)	
1.00	0.99 (0.93 - 1.05)	
2.00	0.68 (0.64 - 0.73)	
5.00	0.40 (0.36 - 0.43)	
10.00	0.28 (0.25 - 0.31)	
20.00	0.21 (0.18 - 0.24)	
50.00	0.13 (0.11 - 0.15)	
Spam Misc%	Ham Misc%	
0.01	96.50 (93.77 - 98.06)	
0.02	90.49 (85.32 - 93.97)	
0.05	75.43 (70.83 - 79.51)	
0.10	63.83 (56.12 - 70.89)	
0.20	21.56 (17.02 - 26.91)	
0.50	3.26 (2.92 - 3.63)	
1.00	0.99 (0.87 - 1.12)	
2.00	0.31 (0.28 - 0.34)	
5.00	0.09 (0.08 - 0.11)	
10.00	0.04 (0.03 - 0.05)	
20.00	0.02 (0.02 - 0.03)	
50.00	0.01 (0.00 - 0.01)	



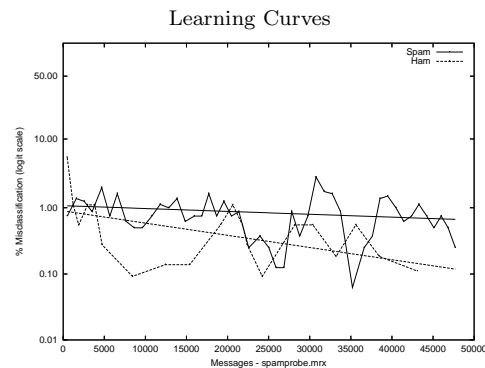
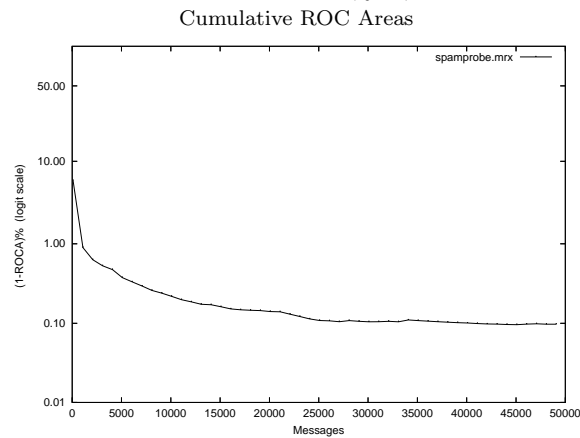
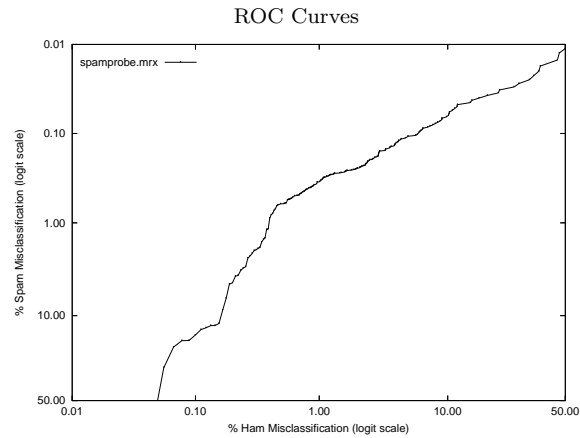
spamprobe - Spamprobe (non-partipant) on Public trec05p-1/full Corpus

Gold Std.	spamprobe.full	
	ham	spam
Filter	39339	1116
Classification	spam	60 51674
Ham Misc%	0.15 (0.12-0.20)	
Spam Misc%	2.11 (1.99-2.24)	
Logit av Misc%	0.57 (0.50-0.65)	
(1-ROCA)%	0.059 (0.049 - 0.071)	
Ham Misc%	Spam Misc%	
0.01	15.30 (2.19 - 59.28)	
0.02	10.37 (6.29 - 16.63)	
0.05	5.68 (4.16 - 7.72)	
0.10	2.77 (1.95 - 3.93)	
0.20	1.76 (1.52 - 2.04)	
0.50	1.06 (0.95 - 1.20)	
1.00	0.65 (0.57 - 0.73)	
2.00	0.41 (0.35 - 0.48)	
5.00	0.21 (0.17 - 0.26)	
10.00	0.11 (0.09 - 0.15)	
20.00	0.05 (0.04 - 0.08)	
50.00	0.01 (0.00 - 0.03)	
Spam Misc%	Ham Misc%	
0.01	52.55 (25.86 - 77.87)	
0.02	39.02 (26.96 - 52.60)	
0.05	20.77 (14.88 - 28.21)	
0.10	12.20 (9.09 - 16.17)	
0.20	5.17 (4.17 - 6.39)	
0.50	1.54 (1.25 - 1.90)	
1.00	0.55 (0.47 - 0.64)	
2.00	0.16 (0.12 - 0.21)	
5.00	0.06 (0.04 - 0.10)	
10.00	0.02 (0.01 - 0.05)	
20.00	0.01 (0.00 - 0.03)	
50.00	0.00 (0.00 - 0.01)	



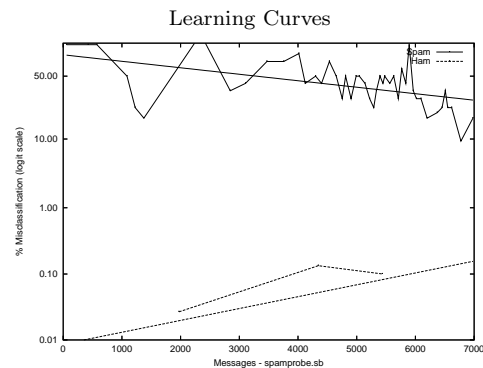
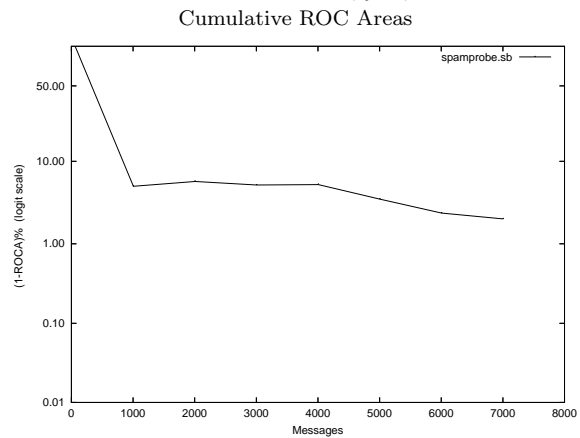
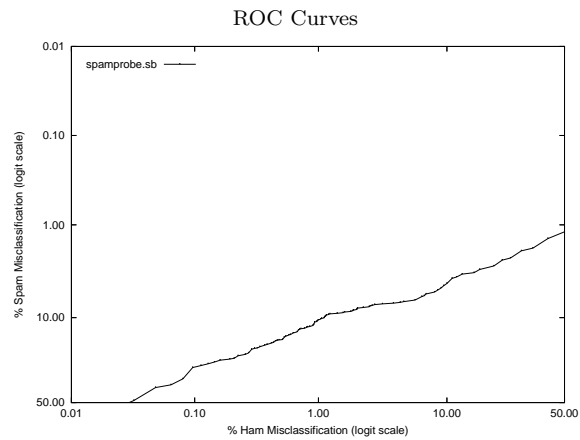
spamprobe - Spamprobe (non-partipant) on Private Mr X. Corpus

Gold Std.	spamprobe.mrx	
	ham	spam
Filter	ham	9001 339
Classification	spam	37 39709
Ham Misc%	0.41 (0.29-0.56)	
Spam Misc%	0.85 (0.76-0.94)	
Logit av Misc%	0.59 (0.50-0.70)	
(1-ROCA)%	0.097 (0.063 - 0.150)	
Ham Misc%	Spam Misc%	
0.01	95.08 (76.24 - 99.15)	
0.02	94.67 (65.21 - 99.41)	
0.05	68.34 (14.92 - 96.37)	
0.10	15.54 (6.58 - 32.46)	
0.20	4.55 (1.61 - 12.18)	
0.50	0.61 (0.38 - 1.00)	
1.00	0.35 (0.27 - 0.46)	
2.00	0.25 (0.20 - 0.31)	
5.00	0.11 (0.08 - 0.16)	
10.00	0.06 (0.04 - 0.10)	
20.00	0.04 (0.02 - 0.06)	
50.00	0.01 (0.00 - 0.04)	
Spam Misc%	Ham Misc%	
0.01	53.04 (36.48 - 68.97)	
0.02	36.10 (20.47 - 55.37)	
0.05	11.64 (6.33 - 20.44)	
0.10	6.10 (3.75 - 9.77)	
0.20	2.53 (1.94 - 3.31)	
0.50	0.63 (0.47 - 0.85)	
1.00	0.40 (0.29 - 0.55)	
2.00	0.31 (0.22 - 0.45)	
5.00	0.19 (0.12 - 0.29)	
10.00	0.17 (0.10 - 0.27)	
20.00	0.08 (0.03 - 0.19)	
50.00	0.06 (0.02 - 0.15)	



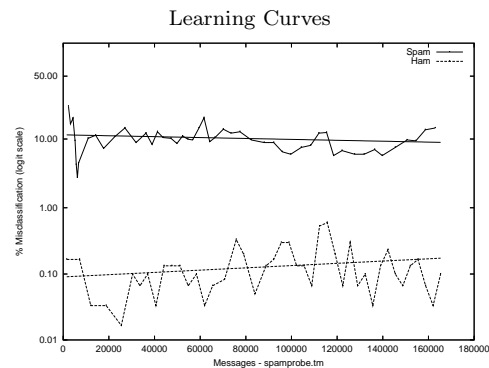
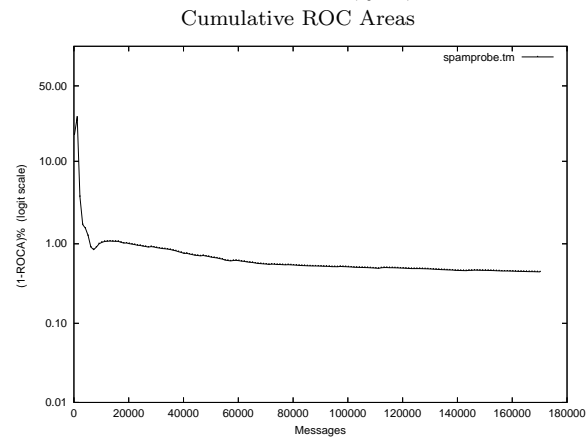
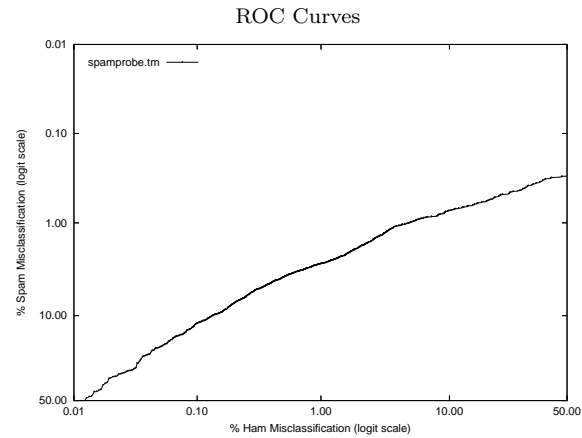
spamprobe - Spamprobe (non-partipant) on Private S.B. Corpus

Gold Std.	spamprobe.sb	
	ham	spam
Filter	ham	6228 326
Classification	spam	3 449
Ham Misc%	0.05 (0.01-0.14)	
Spam Misc%	42.06 (38.56-45.63)	
Logit av Misc%	1.84 (1.05-3.20)	
(1-ROCA)%	2.039 (1.445 - 2.870)	
Ham Misc%	Spam Misc%	
0.01	80.26 (53.38 - 93.52)	
0.02	59.10 (30.30 - 82.76)	
0.05	40.39 (23.11 - 60.43)	
0.10	28.77 (18.81 - 41.32)	
0.20	24.52 (19.68 - 30.09)	
0.50	16.52 (13.04 - 20.70)	
1.00	10.58 (8.27 - 13.44)	
2.00	8.13 (6.51 - 10.11)	
5.00	6.97 (5.43 - 8.90)	
10.00	4.52 (3.06 - 6.62)	
20.00	3.10 (1.95 - 4.88)	
50.00	1.29 (0.60 - 2.73)	
Spam Misc%	Ham Misc%	
0.01	100.00 (nan - nan)	
0.02	100.00 (nan - nan)	
0.05	100.00 (nan - nan)	
0.10	99.37 (nan - nan)	
0.20	98.46 (83.84 - 99.87)	
0.50	69.01 (37.38 - 89.25)	
1.00	61.24 (35.04 - 82.24)	
2.00	29.72 (16.03 - 48.37)	
5.00	8.83 (6.17 - 12.47)	
10.00	1.11 (0.69 - 1.78)	
20.00	0.30 (0.18 - 0.52)	
50.00	0.03 (0.01 - 0.12)	



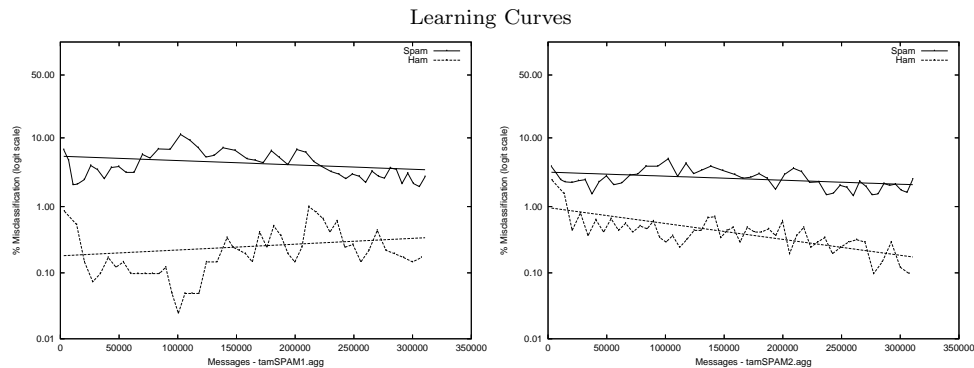
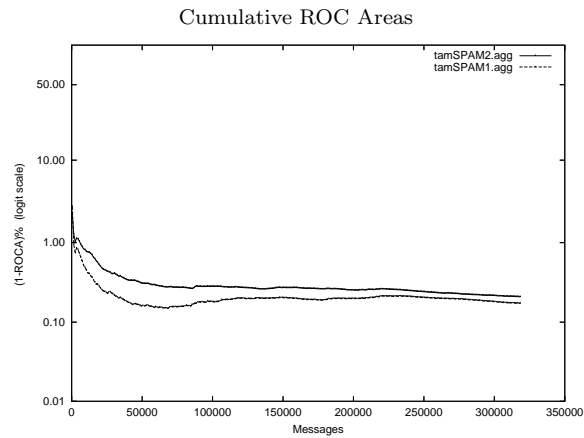
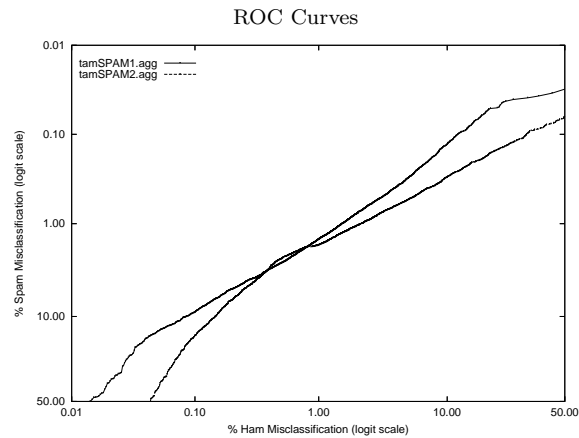
spamprobe - Spamprobe (non-partipant) on Private T.M. Corpus

Gold Std.	spamprobe.tm	
	ham	spam
Filter	150490	2012
Classification	spam	195
Ham Misc%	0.13 (0.11-0.15)	
Spam Misc%	10.31 (9.89-10.74)	
Logit av Misc%	1.21 (1.12-1.30)	
(1-ROCA)%	0.445 (0.385 - 0.513)	
Ham Misc%	Spam Misc%	
0.01	53.45 (34.11 - 71.82)	
0.02	35.77 (27.86 - 44.53)	
0.05	20.12 (17.57 - 22.93)	
0.10	12.02 (10.65 - 13.54)	
0.20	7.45 (6.73 - 8.23)	
0.50	3.97 (3.64 - 4.32)	
1.00	2.82 (2.59 - 3.08)	
2.00	1.89 (1.69 - 2.10)	
5.00	1.00 (0.87 - 1.16)	
10.00	0.73 (0.62 - 0.87)	
20.00	0.54 (0.44 - 0.67)	
50.00	0.30 (0.24 - 0.38)	
Spam Misc%	Ham Misc%	
0.01	99.63 (97.60 - 99.94)	
0.02	99.30 (97.40 - 99.81)	
0.05	96.84 (91.07 - 98.92)	
0.10	86.41 (75.07 - 93.07)	
0.20	73.30 (65.03 - 80.21)	
0.50	21.98 (15.49 - 30.21)	
1.00	5.02 (3.71 - 6.76)	
2.00	1.83 (1.58 - 2.11)	
5.00	0.35 (0.31 - 0.39)	
10.00	0.13 (0.12 - 0.16)	
20.00	0.05 (0.04 - 0.06)	
50.00	0.01 (0.01 - 0.02)	



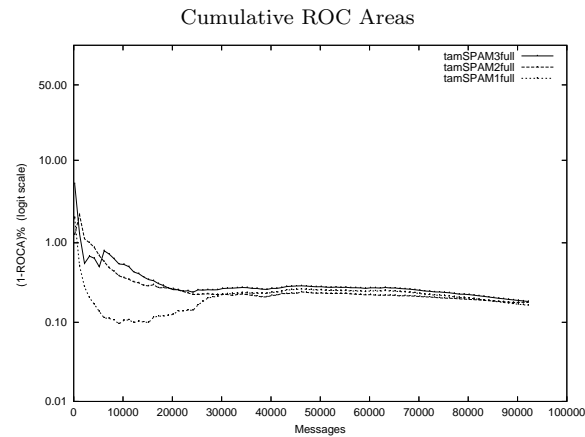
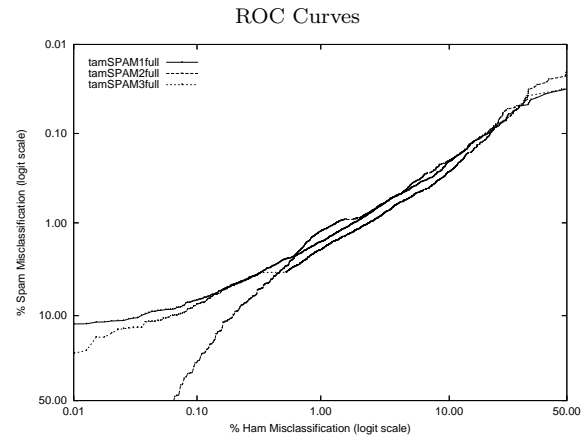
tam - Massey University on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	tamSPAM1.agg		tamSPAM2.agg	
	ham spam	ham spam	ham spam	ham spam
	204834	5006	204420	2974
	519	108123	933	110155
Ham Misc%	0.25 (0.23-0.28)		0.45 (0.43-0.48)	
Spam Misc%	4.43 (4.31-4.55)		2.63 (2.54-2.72)	
Logit av Misc%	1.07 (1.02-1.12)		1.10 (1.06-1.14)	
(1-ROCA)%	0.172 (0.156 - 0.190)		0.209 (0.193 - 0.226)	
Ham Misc%	Spam Misc%			
0.01	77.77 (65.56 - 86.54)		92.43 (88.19 - 95.23)	
0.02	45.53 (27.51 - 64.79)		82.92 (75.41 - 88.49)	
0.05	14.08 (12.42 - 15.91)		42.01 (32.36 - 52.31)	
0.10	9.10 (8.21 - 10.08)		15.50 (13.31 - 17.98)	
0.20	5.22 (4.92 - 5.54)		6.80 (6.23 - 7.40)	
0.50	2.73 (2.60 - 2.86)		2.44 (2.30 - 2.59)	
1.00	1.49 (1.40 - 1.59)		1.72 (1.64 - 1.81)	
2.00	0.78 (0.73 - 0.83)		1.06 (1.00 - 1.13)	
5.00	0.32 (0.29 - 0.35)		0.55 (0.51 - 0.60)	
10.00	0.19 (0.16 - 0.22)		0.30 (0.28 - 0.34)	
20.00	0.16 (0.14 - 0.18)		0.16 (0.14 - 0.19)	
50.00	0.09 (0.08 - 0.12)		0.06 (0.05 - 0.08)	
Spam Misc%	Ham Misc%			
0.01	95.27 (91.40 - 97.45)		96.56 (93.46 - 98.21)	
0.02	90.39 (84.72 - 94.10)		92.73 (86.97 - 96.06)	
0.05	77.05 (70.29 - 82.64)		63.66 (50.51 - 75.04)	
0.10	46.18 (37.22 - 55.39)		32.92 (28.16 - 38.05)	
0.20	7.25 (4.31 - 11.96)		16.32 (13.81 - 19.20)	
0.50	3.25 (2.97 - 3.55)		5.67 (5.15 - 6.25)	
1.00	1.55 (1.45 - 1.65)		2.20 (2.00 - 2.42)	
2.00	0.72 (0.67 - 0.77)		0.68 (0.63 - 0.74)	
5.00	0.21 (0.20 - 0.24)		0.26 (0.24 - 0.28)	
10.00	0.09 (0.07 - 0.10)		0.15 (0.13 - 0.16)	
20.00	0.03 (0.03 - 0.04)		0.08 (0.07 - 0.09)	
50.00	0.02 (0.01 - 0.03)		0.04 (0.04 - 0.05)	

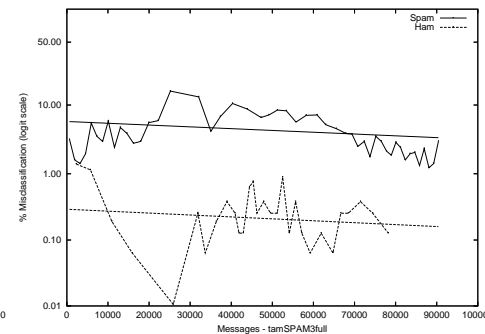
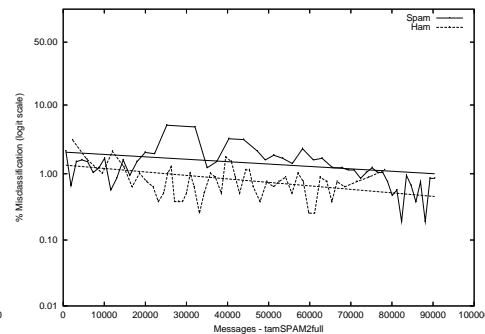
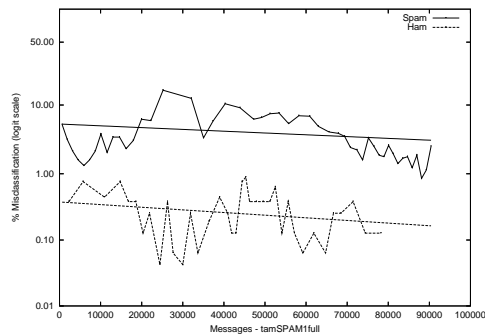


tam - Massey University on Public trec05p-1/full Corpus

Gold Std. Filter Classification	tamSPAM1full			tamSPAM2full			tamSPAM3full		
	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	spam
Ham Misc%	0.26 (0.21-0.32)			0.85 (0.76-0.94)			0.22 (0.18-0.28)		
Spam Misc%	4.10 (3.93-4.27)			1.45 (1.35-1.55)			4.46 (4.28-4.64)		
Logit av Misc%	1.05 (0.95-1.15)			1.11 (1.04-1.18)			1.01 (0.91-1.12)		
(1-ROCA)%	0.164 (0.142 - 0.189)			0.178 (0.152 - 0.207)			0.183 (0.163 - 0.205)		
Ham Misc%	Spam Misc%								
0.01	12.55 (9.59 - 16.25)			97.30 (90.63 - 99.26)			22.94 (4.64 - 64.58)		
0.02	11.57 (10.43 - 12.83)			89.67 (71.91 - 96.72)			16.24 (10.32 - 24.61)		
0.05	8.87 (7.81 - 10.06)			59.90 (38.67 - 77.97)			11.24 (9.39 - 13.41)		
0.10	6.92 (5.96 - 8.02)			27.38 (16.89 - 41.15)			7.64 (6.15 - 9.45)		
0.20	4.73 (4.13 - 5.42)			9.90 (7.25 - 13.38)			4.86 (4.16 - 5.67)		
0.50	2.58 (2.32 - 2.87)			3.21 (2.66 - 3.87)			3.49 (3.29 - 3.71)		
1.00	1.63 (1.49 - 1.79)			1.25 (1.11 - 1.41)			2.00 (1.80 - 2.22)		
2.00	0.94 (0.84 - 1.05)			0.89 (0.80 - 0.98)			1.18 (1.06 - 1.31)		
5.00	0.43 (0.37 - 0.50)			0.40 (0.34 - 0.46)			0.54 (0.48 - 0.60)		
10.00	0.21 (0.17 - 0.25)			0.20 (0.17 - 0.25)			0.27 (0.23 - 0.32)		
20.00	0.15 (0.12 - 0.19)			0.09 (0.07 - 0.12)			0.16 (0.13 - 0.20)		
50.00	0.09 (0.07 - 0.11)			0.02 (0.01 - 0.04)			0.09 (0.07 - 0.11)		
Spam Misc%	Ham Misc%								
0.01	96.83 (92.67 - 98.67)			91.33 (52.16 - 99.03)			98.04 (95.49 - 99.16)		
0.02	93.55 (87.83 - 96.68)			51.73 (24.23 - 78.22)			93.34 (87.17 - 96.66)		
0.05	79.36 (70.17 - 86.28)			28.22 (20.26 - 37.83)			75.56 (66.89 - 82.55)		
0.10	45.79 (32.52 - 59.69)			18.89 (15.07 - 23.41)			43.69 (31.09 - 57.16)		
0.20	10.57 (8.41 - 13.20)			10.13 (8.21 - 12.43)			12.31 (9.96 - 15.12)		
0.50	4.06 (3.35 - 4.91)			3.94 (3.34 - 4.66)			5.40 (4.61 - 6.32)		
1.00	1.84 (1.59 - 2.13)			1.34 (1.15 - 1.56)			2.46 (2.16 - 2.81)		
2.00	0.77 (0.67 - 0.88)			0.69 (0.60 - 0.79)			1.00 (0.88 - 1.15)		
5.00	0.19 (0.15 - 0.23)			0.35 (0.29 - 0.42)			0.19 (0.14 - 0.25)		
10.00	0.04 (0.02 - 0.06)			0.20 (0.15 - 0.25)			0.06 (0.04 - 0.09)		
20.00	0.00 (0.00 - 0.01)			0.13 (0.09 - 0.17)			0.02 (0.01 - 0.04)		
50.00	0.00 (0.00 - 0.00)			0.07 (0.04 - 0.11)			0.01 (0.00 - 0.02)		

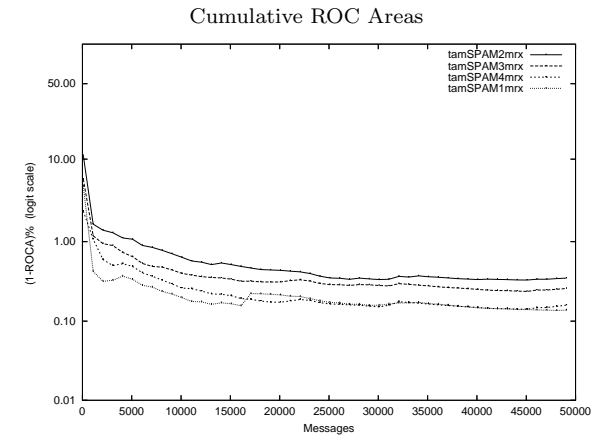
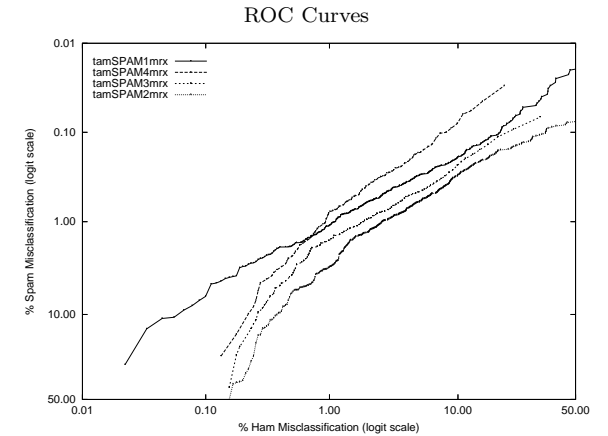


Learning Curves

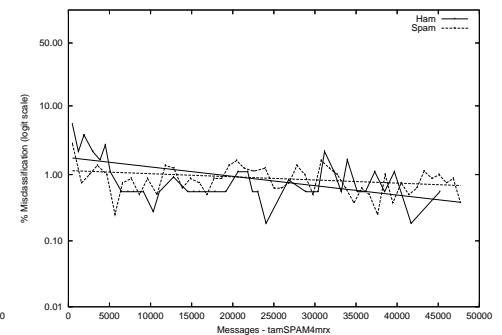
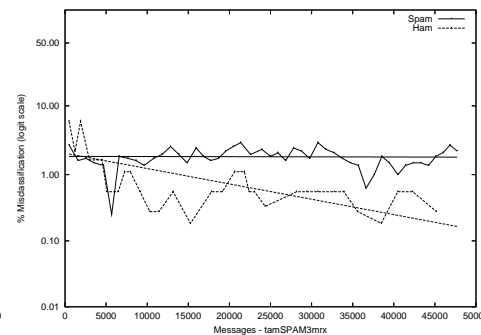
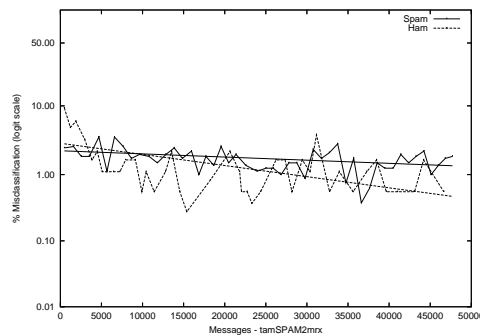
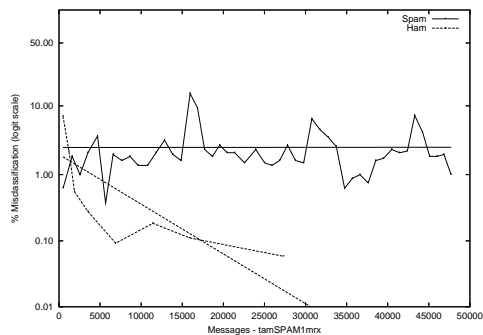


tam - Massey University on Private Mr X. Corpus

Gold Std. Filter Classification	tamSPAM1mrx		tamSPAM2mrx		tamSPAM3mrx		tamSPAM4mrx	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.28 (0.18-0.41)		1.43 (1.19-1.69)		0.82 (0.64-1.03)		0.96 (0.77-1.19)	
Spam Misc%	2.55 (2.39-2.71)		1.75 (1.63-1.89)		1.85 (1.72-1.98)		0.89 (0.80-0.98)	
Logit av Misc%	0.84 (0.69-1.03)		1.58 (1.44-1.74)		1.23 (1.09-1.39)		0.92 (0.82-1.04)	
(1-ROCA)%	0.138 (0.111 - 0.171)		0.349 (0.285 - 0.426)		0.257 (0.195 - 0.340)		0.159 (0.110 - 0.229)	
Ham Misc%	Spam Misc%							
0.01	45.65 (12.84 - 82.73)	98.21 (90.76 - 99.67)	97.03 (78.07 - 99.67)	90.01 (63.14 - 97.93)				
0.02	40.96 (11.14 - 79.34)	97.44 (90.41 - 99.36)	89.09 (65.35 - 97.25)	85.85 (59.75 - 96.12)				
0.05	10.97 (3.17 - 31.66)	91.09 (74.40 - 97.29)	84.26 (57.27 - 95.53)	79.43 (47.59 - 94.26)				
0.10	6.51 (3.01 - 13.51)	78.36 (44.65 - 94.20)	58.80 (28.22 - 83.82)	46.24 (14.81 - 80.98)				
0.20	3.12 (2.07 - 4.67)	38.08 (15.33 - 67.63)	20.17 (5.61 - 51.80)	13.48 (4.37 - 34.66)				
0.50	1.86 (1.54 - 2.23)	6.00 (3.88 - 9.17)	3.72 (2.43 - 5.67)	2.38 (1.63 - 3.47)				
1.00	1.10 (0.88 - 1.39)	3.17 (2.50 - 4.01)	1.57 (1.34 - 1.85)	0.77 (0.57 - 1.05)				
2.00	0.61 (0.50 - 0.74)	1.33 (1.15 - 1.53)	0.94 (0.80 - 1.10)	0.40 (0.32 - 0.49)				
5.00	0.31 (0.26 - 0.38)	0.61 (0.54 - 0.69)	0.47 (0.40 - 0.55)	0.17 (0.13 - 0.22)				
10.00	0.19 (0.15 - 0.24)	0.30 (0.26 - 0.36)	0.23 (0.19 - 0.29)	0.07 (0.05 - 0.11)				
20.00	0.12 (0.09 - 0.16)	0.16 (0.12 - 0.20)	0.11 (0.08 - 0.15)	0.03 (0.02 - 0.06)				
50.00	0.08 (0.06 - 0.11)	0.08 (0.05 - 0.11)	0.05 (0.03 - 0.08)	0.02 (0.01 - 0.05)				
Spam Misc%	Ham Misc%							
0.01	90.76 (78.75 - 96.30)	86.01 (73.71 - 93.10)	88.17 (70.56 - 95.87)	78.26 (45.29 - 93.99)				
0.02	87.76 (76.43 - 94.07)	83.37 (73.49 - 90.07)	76.48 (58.99 - 88.02)	48.97 (18.58 - 80.15)				
0.05	74.12 (59.53 - 84.80)	65.06 (51.19 - 76.77)	55.81 (35.81 - 74.09)	13.20 (8.63 - 19.66)				
0.10	30.90 (13.62 - 55.92)	36.31 (26.55 - 47.35)	21.86 (14.27 - 32.00)	7.55 (5.59 - 10.11)				
0.20	9.67 (7.66 - 12.14)	16.00 (12.07 - 20.91)	11.14 (9.11 - 13.55)	4.40 (3.52 - 5.50)				
0.50	2.64 (2.12 - 3.30)	6.28 (5.29 - 7.45)	4.39 (3.49 - 5.52)	1.71 (1.38 - 2.13)				
1.00	1.12 (0.90 - 1.39)	2.63 (2.24 - 3.09)	1.89 (1.58 - 2.26)	0.93 (0.77 - 1.12)				
2.00	0.38 (0.25 - 0.56)	1.31 (1.08 - 1.57)	0.72 (0.57 - 0.90)	0.56 (0.44 - 0.72)				
5.00	0.11 (0.06 - 0.21)	0.66 (0.50 - 0.89)	0.40 (0.28 - 0.57)	0.28 (0.19 - 0.41)				
10.00	0.07 (0.03 - 0.17)	0.39 (0.28 - 0.53)	0.27 (0.17 - 0.41)	0.22 (0.14 - 0.35)				
20.00	0.03 (0.01 - 0.16)	0.25 (0.17 - 0.37)	0.22 (0.14 - 0.36)	0.18 (0.10 - 0.31)				
50.00	0.01 (0.00 - 0.02)	0.17 (0.10 - 0.28)	0.11 (0.06 - 0.21)	0.10 (0.04 - 0.23)				

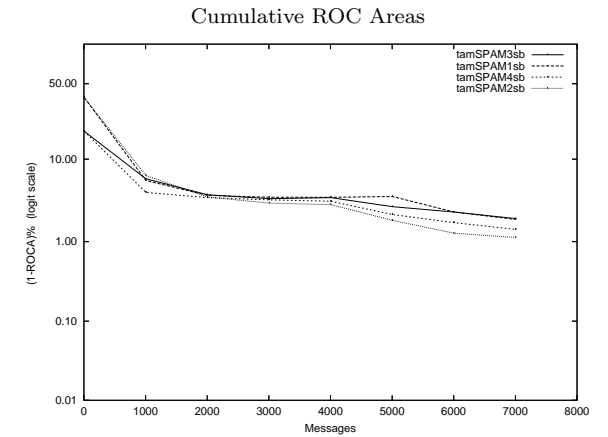
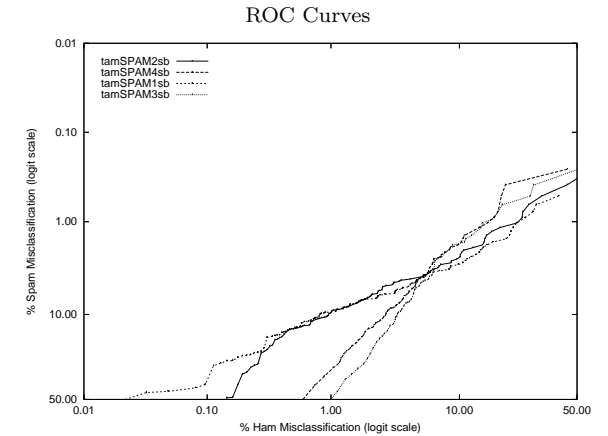


Learning Curves

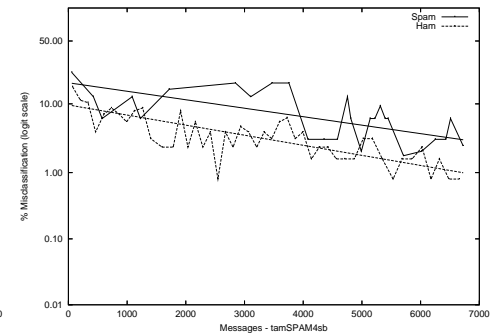
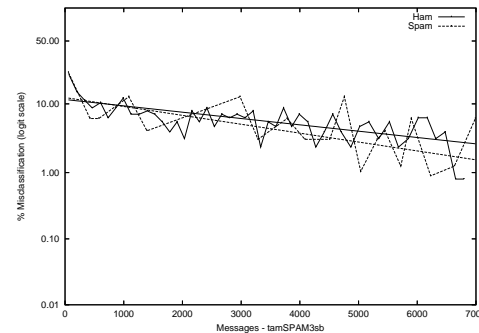
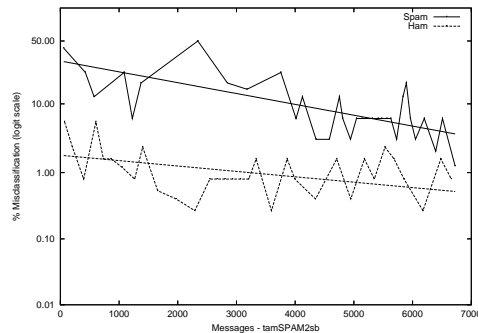
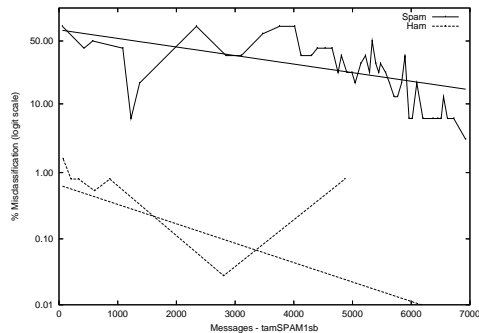


tam - Massey University on Private S.B. Corpus

Gold Std. Filter Classification	tamSPAM1sb		tamSPAM2sb		tamSPAM3sb		tamSPAM4sb	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.14 (0.07-0.27)		1.04 (0.81-1.33)		6.29 (5.70-6.92)		3.92 (3.45-4.43)	
Spam Misc%	27.48 (24.37-30.77)		9.29 (7.34-11.56)		3.64 (2.43-5.21)		6.19 (4.60-8.13)	
Logit av Misc%	2.29 (1.65-3.17)		3.18 (2.69-3.76)		4.79 (3.97-5.76)		4.93 (4.23-5.73)	
(1-ROCA)%	1.892 (1.321 - 2.701)		1.127 (0.757 - 1.673)		1.934 (1.539 - 2.429)		1.421 (1.140 - 1.772)	
Ham Misc%	Spam Misc%							
0.01	53.81 (47.29 - 60.19)	88.77 (84.05 - 92.23)	99.35 (nan - nan)	98.71 (nan - nan)	98.71 (nan - nan)	98.71 (nan - nan)	98.71 (nan - nan)	98.71 (nan - nan)
0.02	53.42 (43.15 - 63.40)	84.90 (74.49 - 91.55)	98.57 (nan - nan)	97.16 (nan - nan)	98.57 (nan - nan)	97.16 (nan - nan)	97.16 (nan - nan)	97.16 (nan - nan)
0.05	44.77 (32.37 - 57.86)	84.00 (62.41 - 94.32)	98.57 (nan - nan)	96.13 (nan - nan)	98.57 (nan - nan)	96.13 (nan - nan)	96.13 (nan - nan)	96.13 (nan - nan)
0.10	40.52 (24.81 - 58.43)	66.06 (32.21 - 88.86)	96.49 (92.05 - 98.49)	89.68 (77.23 - 95.70)	96.49 (92.05 - 98.49)	89.68 (77.23 - 95.70)	89.68 (77.23 - 95.70)	89.68 (77.23 - 95.70)
0.20	24.52 (17.82 - 32.73)	34.19 (16.45 - 57.83)	91.43 (83.46 - 95.75)	81.94 (69.78 - 89.91)	91.43 (83.46 - 95.75)	81.94 (69.78 - 89.91)	81.94 (69.78 - 89.91)	81.94 (69.78 - 89.91)
0.50	13.94 (10.50 - 18.27)	13.94 (10.40 - 18.42)	74.55 (61.60 - 84.24)	62.06 (44.96 - 76.62)	74.55 (61.60 - 84.24)	62.06 (44.96 - 76.62)	62.06 (44.96 - 76.62)	62.06 (44.96 - 76.62)
1.00	9.55 (7.62 - 11.89)	9.68 (7.21 - 12.87)	50.13 (39.62 - 60.63)	31.87 (23.28 - 41.90)	50.13 (39.62 - 60.63)	31.87 (23.28 - 41.90)	31.87 (23.28 - 41.90)	31.87 (23.28 - 41.90)
2.00	6.97 (5.58 - 8.67)	6.71 (5.07 - 8.83)	24.81 (18.14 - 32.94)	13.81 (10.48 - 17.97)	24.81 (18.14 - 32.94)	13.81 (10.48 - 17.97)	13.81 (10.48 - 17.97)	13.81 (10.48 - 17.97)
5.00	4.26 (3.10 - 5.82)	4.26 (2.89 - 6.23)	5.06 (3.63 - 7.02)	4.13 (2.82 - 6.01)	5.06 (3.63 - 7.02)	4.13 (2.82 - 6.01)	4.13 (2.82 - 6.01)	4.13 (2.82 - 6.01)
10.00	3.10 (2.17 - 4.40)	2.58 (1.53 - 4.31)	1.82 (1.03 - 3.20)	1.81 (1.08 - 3.02)	1.82 (1.03 - 3.20)	1.81 (1.08 - 3.02)	1.81 (1.08 - 3.02)	1.81 (1.08 - 3.02)
20.00	2.32 (1.53 - 3.52)	1.16 (0.56 - 2.41)	0.65 (0.23 - 1.78)	0.39 (0.10 - 1.42)	0.65 (0.23 - 1.78)	0.39 (0.10 - 1.42)	0.39 (0.10 - 1.42)	0.39 (0.10 - 1.42)
50.00	1.68 (0.97 - 2.90)	0.39 (0.11 - 1.34)	0.26 (0.06 - 1.04)	0.26 (0.07 - 0.98)	0.26 (0.06 - 1.04)	0.26 (0.07 - 0.98)	0.26 (0.07 - 0.98)	0.26 (0.07 - 0.98)
Spam Misc%	Ham Misc%							
0.01	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)
0.02	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)
0.05	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)	100.00 (nan - nan)
0.10	99.25 (nan - nan)	93.47 (72.01 - 98.76)	96.41 (50.96 - 99.86)	58.39 (23.06 - 86.78)	96.41 (50.96 - 99.86)	58.39 (23.06 - 86.78)	58.39 (23.06 - 86.78)	58.39 (23.06 - 86.78)
0.20	97.90 (80.15 - 99.81)	90.50 (43.46 - 99.16)	78.56 (20.33 - 98.13)	54.16 (20.27 - 84.60)	78.56 (20.33 - 98.13)	54.16 (20.27 - 84.60)	54.16 (20.27 - 84.60)	54.16 (20.27 - 84.60)
0.50	89.10 (66.92 - 97.06)	45.43 (12.68 - 82.68)	30.79 (6.41 - 74.30)	19.74 (6.94 - 44.77)	30.79 (6.41 - 74.30)	19.74 (6.94 - 44.77)	19.74 (6.94 - 44.77)	19.74 (6.94 - 44.77)
1.00	74.05 (45.02 - 90.86)	26.38 (13.42 - 45.32)	17.43 (9.19 - 30.57)	16.61 (10.22 - 25.85)	17.43 (9.19 - 30.57)	16.61 (10.22 - 25.85)	16.61 (10.22 - 25.85)	16.61 (10.22 - 25.85)
2.00	30.25 (8.43 - 67.14)	12.52 (6.94 - 21.55)	8.32 (5.20 - 13.06)	8.57 (5.52 - 13.08)	8.32 (5.20 - 13.06)	8.57 (5.52 - 13.08)	8.57 (5.52 - 13.08)	8.57 (5.52 - 13.08)
5.00	4.25 (2.81 - 6.39)	3.08 (1.57 - 5.97)	5.07 (3.99 - 6.42)	4.67 (3.64 - 5.98)	5.07 (3.99 - 6.42)	4.67 (3.64 - 5.98)	4.67 (3.64 - 5.98)	4.67 (3.64 - 5.98)
10.00	0.82 (0.51 - 1.32)	0.96 (0.59 - 1.56)	3.35 (2.82 - 3.98)	2.74 (2.18 - 3.45)	3.35 (2.82 - 3.98)	2.74 (2.18 - 3.45)	2.74 (2.18 - 3.45)	2.74 (2.18 - 3.45)
20.00	0.30 (0.18 - 0.51)	0.34 (0.21 - 0.54)	2.25 (1.89 - 2.66)	1.48 (1.13 - 1.92)	2.25 (1.89 - 2.66)	1.48 (1.13 - 1.92)	1.48 (1.13 - 1.92)	1.48 (1.13 - 1.92)
50.00	0.03 (0.01 - 0.12)	0.14 (0.07 - 0.31)	1.01 (0.77 - 1.33)	0.61 (0.43 - 0.86)	1.01 (0.77 - 1.33)	0.61 (0.43 - 0.86)	0.61 (0.43 - 0.86)	0.61 (0.43 - 0.86)

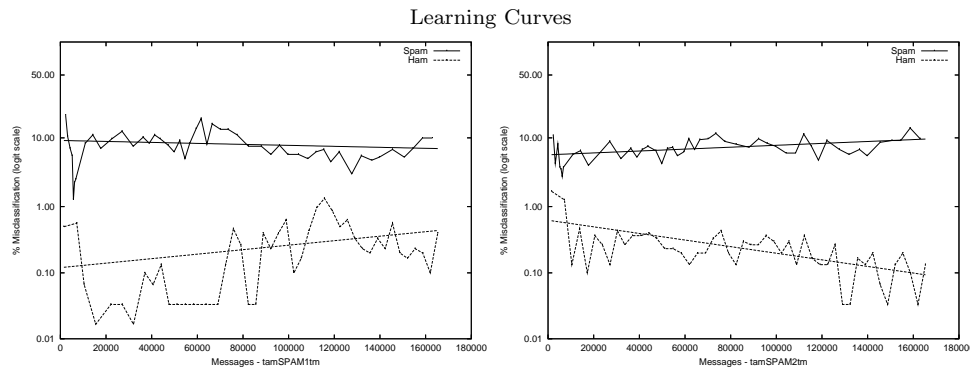
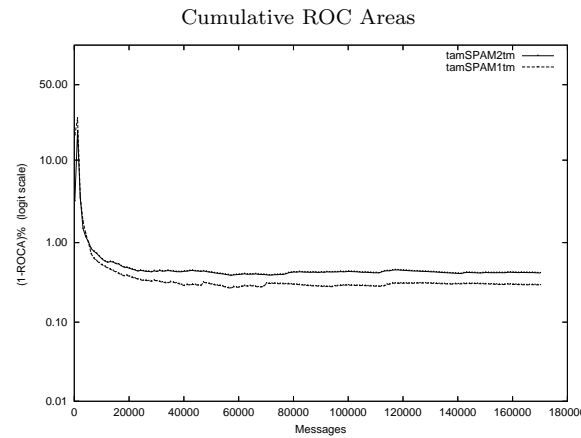
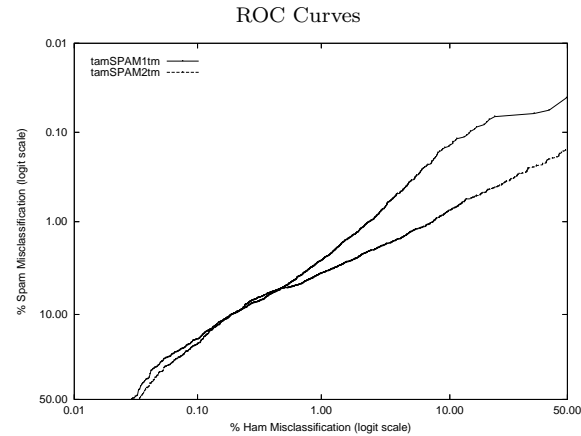


Learning Curves



tam - Massey University on Private T.M. Corpus

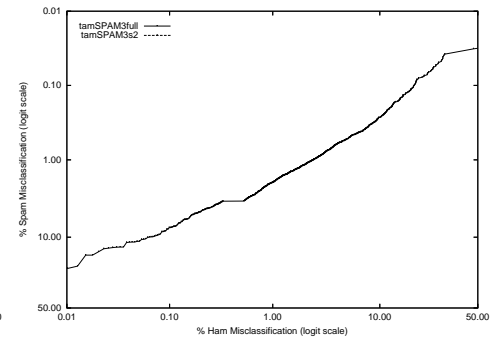
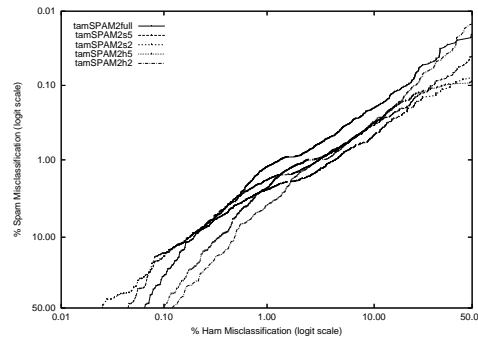
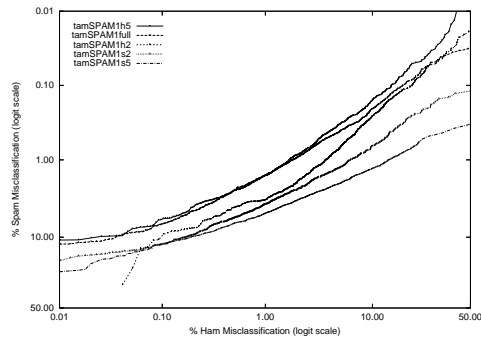
Gold Std. Filter Classification	tamSPAM1tm		tamSPAM2tm	
	ham spam	ham spam	ham spam	ham spam
	150303	1611	150280	1437
	382	17905	405	18079
Ham Misc%	0.25 (0.23-0.28)		0.27 (0.24-0.30)	
Spam Misc%	8.25 (7.87-8.65)		7.36 (7.00-7.74)	
Logit av Misc%	1.49 (1.41-1.57)		1.44 (1.37-1.52)	
(1-ROCA)%	0.294 (0.244 - 0.354)		0.416 (0.363 - 0.477)	
Ham Misc%	Spam Misc%			
0.01	82.25 (71.71 - 89.43)	89.02 (80.51 - 94.09)	89.02 (80.51 - 94.09)	82.25 (71.71 - 89.43)
0.02	65.04 (51.67 - 76.41)	73.93 (54.41 - 87.08)	73.93 (54.41 - 87.08)	65.04 (51.67 - 76.41)
0.05	27.98 (22.70 - 33.94)	32.99 (25.76 - 41.14)	32.99 (25.76 - 41.14)	27.98 (22.70 - 33.94)
0.10	17.40 (15.59 - 19.37)	19.41 (16.56 - 22.63)	19.41 (16.56 - 22.63)	17.40 (15.59 - 19.37)
0.20	9.68 (8.95 - 10.46)	9.79 (9.01 - 10.63)	9.79 (9.01 - 10.63)	9.68 (8.95 - 10.46)
0.50	5.21 (4.76 - 5.69)	5.33 (5.04 - 5.63)	5.33 (5.04 - 5.63)	5.21 (4.76 - 5.69)
1.00	2.70 (2.45 - 2.97)	3.70 (3.42 - 4.00)	3.70 (3.42 - 4.00)	2.70 (2.45 - 2.97)
2.00	1.27 (1.11 - 1.46)	2.44 (2.20 - 2.72)	2.44 (2.20 - 2.72)	1.27 (1.11 - 1.46)
5.00	0.36 (0.28 - 0.47)	1.33 (1.16 - 1.51)	1.33 (1.16 - 1.51)	0.36 (0.28 - 0.47)
10.00	0.31 (0.24 - 0.41)	0.75 (0.64 - 0.89)	0.75 (0.64 - 0.89)	0.31 (0.24 - 0.41)
20.00	0.29 (0.22 - 0.39)	0.43 (0.35 - 0.52)	0.43 (0.35 - 0.52)	0.29 (0.22 - 0.39)
50.00	0.18 (0.13 - 0.27)	0.16 (0.11 - 0.23)	0.16 (0.11 - 0.23)	0.18 (0.13 - 0.27)
Spam Misc%	Ham Misc%			
0.01	97.07 (90.58 - 99.13)	99.58 (97.79 - 99.92)	99.58 (97.79 - 99.92)	97.07 (90.58 - 99.13)
0.02	93.75 (85.75 - 97.40)	99.06 (95.24 - 99.82)	99.06 (95.24 - 99.82)	93.75 (85.75 - 97.40)
0.05	83.06 (71.39 - 90.60)	91.41 (80.18 - 96.55)	91.41 (80.18 - 96.55)	83.06 (71.39 - 90.60)
0.10	74.30 (59.46 - 85.06)	75.70 (52.21 - 89.88)	75.70 (52.21 - 89.88)	74.30 (59.46 - 85.06)
0.20	40.94 (20.33 - 65.31)	41.44 (30.38 - 53.45)	41.44 (30.38 - 53.45)	40.94 (20.33 - 65.31)
0.50	3.95 (3.33 - 4.66)	16.46 (12.65 - 21.14)	16.46 (12.65 - 21.14)	3.95 (3.33 - 4.66)
1.00	2.42 (2.10 - 2.78)	7.57 (6.45 - 8.87)	7.57 (6.45 - 8.87)	2.42 (2.10 - 2.78)
2.00	1.31 (1.19 - 1.44)	2.74 (2.30 - 3.26)	2.74 (2.30 - 3.26)	1.31 (1.19 - 1.44)
5.00	0.52 (0.47 - 0.58)	0.62 (0.53 - 0.73)	0.62 (0.53 - 0.73)	0.52 (0.47 - 0.58)
10.00	0.19 (0.17 - 0.22)	0.19 (0.17 - 0.22)	0.19 (0.17 - 0.22)	0.19 (0.17 - 0.22)
20.00	0.08 (0.07 - 0.09)	0.09 (0.08 - 0.11)	0.09 (0.08 - 0.11)	0.08 (0.07 - 0.09)
50.00	0.03 (0.02 - 0.04)	0.03 (0.03 - 0.05)	0.03 (0.03 - 0.05)	0.03 (0.02 - 0.04)



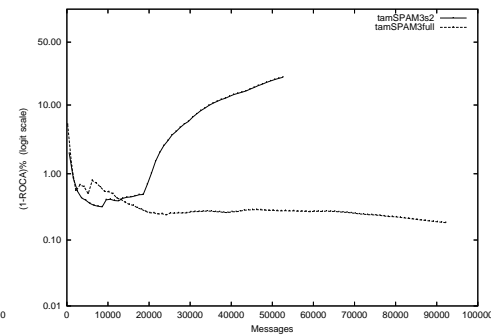
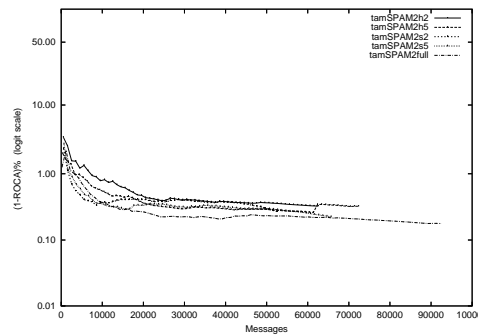
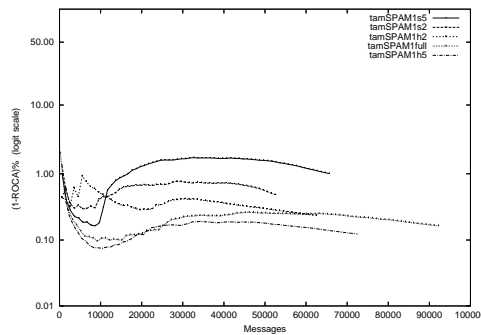
tam - Massey University - Public subset Corpora

	tamSPAM1full	tamSPAM1h2	tamSPAM1h5	tamSPAM1s2
Ham Misc%	0.26 (0.21-0.32)	0.47 (0.35-0.63)	0.37 (0.29-0.46)	0.22 (0.17-0.27)
Spam Misc%	4.10 (3.93-4.27)	4.55 (4.37-4.73)	3.15 (3.00-3.30)	9.05 (8.57-9.56)
Logit av Misc%	1.05 (0.95-1.15)	1.48 (1.28-1.71)	1.08 (0.96-1.22)	1.45 (1.30-1.62)
(1-ROCA)%	0.164 (0.142 - 0.189)	0.234 (0.196 - 0.281)	0.123 (0.105 - 0.144)	0.483 (0.401 - 0.582)
	tamSPAM1s5	tamSPAM2full	tamSPAM2h2	tamSPAM2h5
Ham Misc%	0.07 (0.05-0.11)	0.85 (0.76-0.94)	1.97 (1.70-2.26)	1.42 (1.26-1.60)
Spam Misc%	13.94 (13.52-14.37)	1.45 (1.35-1.55)	1.56 (1.46-1.67)	1.51 (1.41-1.62)
Logit av Misc%	1.08 (0.90-1.29)	1.11 (1.04-1.18)	1.75 (1.62-1.89)	1.46 (1.37-1.57)
(1-ROCA)%	1.004 (0.916 - 1.100)	0.178 (0.152 - 0.207)	0.326 (0.264 - 0.402)	0.323 (0.276 - 0.379)
	tamSPAM2s2	tamSPAM2s5	tamSPAM3full	tamSPAM3s2
Ham Misc%	0.73 (0.65-0.82)	0.72 (0.63-0.80)	0.22 (0.18-0.28)	0.34 (0.29-0.40)
Spam Misc%	3.03 (2.74-3.33)	2.39 (2.21-2.59)	4.46 (4.28-4.64)	69.17 (68.37-69.96)
Logit av Misc%	1.49 (1.39-1.61)	1.31 (1.22-1.41)	1.01 (0.91-1.12)	8.05 (7.43-8.71)
(1-ROCA)%	0.268 (0.224 - 0.320)	0.225 (0.195 - 0.260)	0.183 (0.163 - 0.205)	22.663 (22.258 - 23.073)

ROC Curves

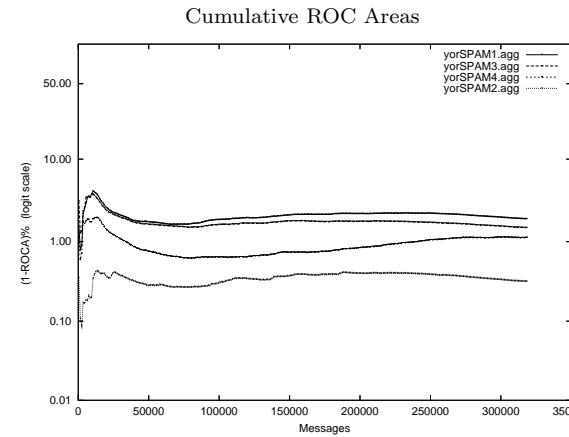
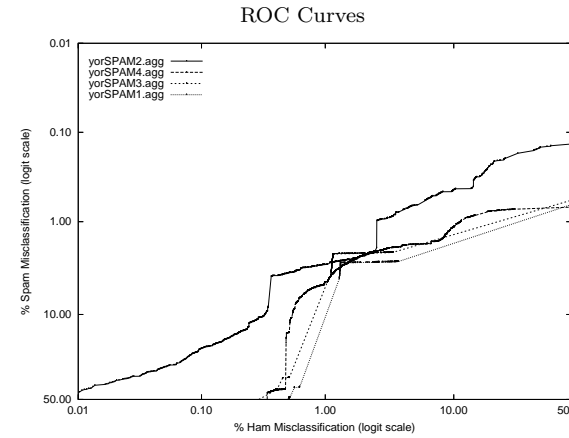


Cumulative ROC Area

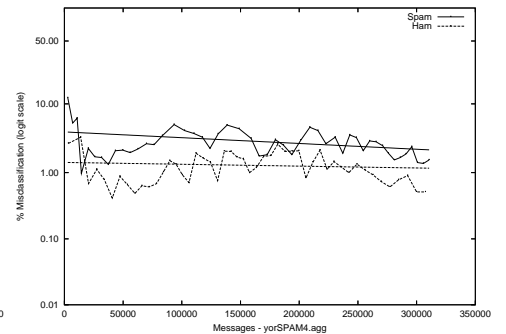
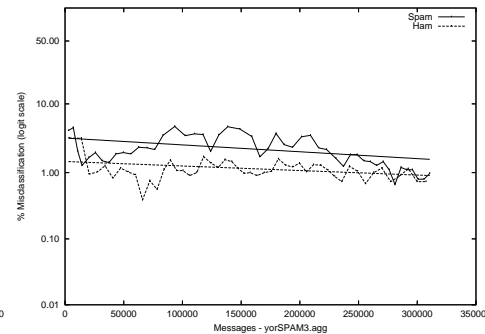
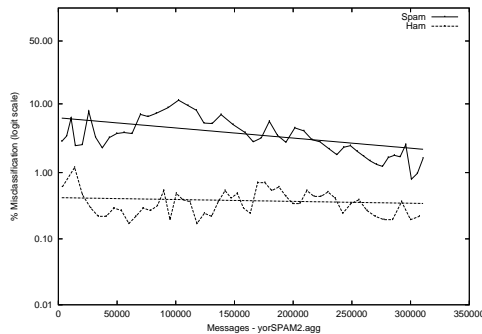
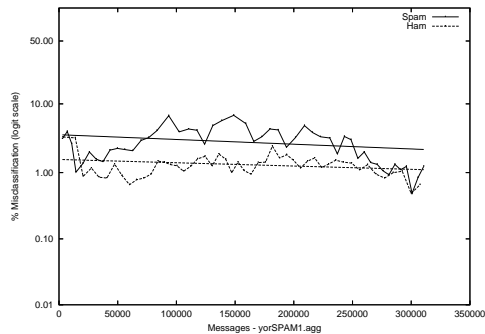


yor - York U. on Aggregate of full, mrx, sb, tm Corpus

Gold Std. Filter Classification	yorSPAM1.agg		yorSPAM2.agg		yorSPAM3.agg		yorSPAM4.agg	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	1.32 (1.27-1.37)		0.38 (0.35-0.41)		1.16 (1.11-1.21)		1.29 (1.24-1.34)	
Spam Misc%	2.85 (2.75-2.95)		3.91 (3.80-4.02)		2.29 (2.21-2.38)		2.98 (2.88-3.08)	
Logit av Misc%	1.94 (1.90-1.99)		1.23 (1.18-1.28)		1.63 (1.59-1.68)		1.96 (1.91-2.01)	
(1-ROCA)%	1.917 (1.856 - 1.981)		0.316 (0.294 - 0.339)		1.491 (1.443 - 1.540)		1.122 (1.073 - 1.174)	
Ham Misc%	Spam Misc%							
0.01	95.73 (94.72 - 96.56)	45.69 (39.81 - 51.68)	88.46 (86.51 - 90.15)	96.07 (95.33 - 96.69)	93.66 (92.57 - 94.60)	38.44 (35.81 - 41.13)	85.55 (83.49 - 87.40)	94.60 (93.74 - 95.35)
0.02	89.95 (88.77 - 91.02)	30.41 (28.58 - 32.29)	78.42 (76.98 - 79.79)	89.26 (87.48 - 90.80)	84.38 (83.24 - 85.45)	21.14 (19.67 - 22.68)	70.88 (68.45 - 73.19)	81.80 (80.27 - 83.24)
0.05	73.95 (72.11 - 75.70)	15.13 (14.10 - 16.22)	60.10 (56.49 - 63.60)	68.38 (67.64 - 69.10)	49.81 (47.76 - 51.87)	3.67 (3.53 - 3.82)	36.45 (35.62 - 37.29)	15.22 (5.92 - 33.88)
0.10	20.87 (18.82 - 23.09)	2.96 (2.86 - 3.08)	10.33 (8.11 - 13.08)	4.60 (4.31 - 4.92)	2.82 (2.72 - 2.92)	2.40 (2.31 - 2.49)	2.24 (2.17 - 2.31)	2.37 (2.26 - 2.49)
0.20	2.71 (2.61 - 2.81)	0.66 (0.62 - 0.71)	2.14 (2.06 - 2.22)	1.80 (1.70 - 1.89)	2.55 (2.45 - 2.65)	0.44 (0.41 - 0.48)	2.02 (1.95 - 2.10)	1.15 (1.08 - 1.22)
0.50	2.23 (2.14 - 2.32)	0.21 (0.18 - 0.24)	1.78 (1.71 - 1.86)	0.76 (0.71 - 0.80)	1.00	0.13 (0.11 - 0.16)	1.09 (1.04 - 1.15)	0.68 (0.63 - 0.72)
1.00	1.38 (1.32 - 1.46)	0.13 (0.11 - 0.16)	1.09 (1.04 - 1.15)	0.68 (0.63 - 0.72)	5.00	2.71 (2.61 - 2.81)	0.66 (0.62 - 0.71)	2.14 (2.06 - 2.22)
2.00	2.71 (2.61 - 2.81)	0.66 (0.62 - 0.71)	2.14 (2.06 - 2.22)	1.80 (1.70 - 1.89)	10.00	2.55 (2.45 - 2.65)	0.44 (0.41 - 0.48)	2.02 (1.95 - 2.10)
5.00	2.23 (2.14 - 2.32)	0.21 (0.18 - 0.24)	1.78 (1.71 - 1.86)	0.76 (0.71 - 0.80)	20.00	2.23 (2.14 - 2.32)	0.21 (0.18 - 0.24)	1.78 (1.71 - 1.86)
10.00	1.38 (1.32 - 1.46)	0.13 (0.11 - 0.16)	1.09 (1.04 - 1.15)	0.68 (0.63 - 0.72)	50.00	1.38 (1.32 - 1.46)	0.13 (0.11 - 0.16)	1.09 (1.04 - 1.15)
Spam Misc%	Ham Misc%							
0.01	98.15 (96.85 - 98.92)	99.94 (99.91 - 99.96)	97.52 (96.45 - 98.27)	99.96 (99.92 - 99.98)	97.33 (96.82 - 97.75)	99.93 (99.92 - 99.94)	96.63 (95.98 - 97.18)	99.93 (99.89 - 99.96)
0.02	95.67 (95.08 - 96.20)	99.93 (99.92 - 99.94)	94.80 (93.99 - 95.51)	99.83 (99.77 - 99.88)	94.16 (93.62 - 94.66)	99.93 (94.04 - 100.00)	92.89 (92.23 - 93.50)	99.66 (99.59 - 99.73)
0.05	91.40 (90.61 - 92.13)	22.22 (17.91 - 27.22)	89.11 (88.10 - 90.05)	99.26 (99.15 - 99.36)	81.00 (79.42 - 82.48)	7.58 (6.58 - 8.71)	76.62 (74.70 - 78.43)	98.20 (98.03 - 98.35)
0.10	64.18 (62.09 - 66.22)	2.57 (2.44 - 2.71)	54.84 (52.49 - 57.16)	11.21 (10.60 - 11.85)	64.18 (62.09 - 66.22)	2.57 (2.44 - 2.71)	54.84 (52.49 - 57.16)	11.21 (10.60 - 11.85)
0.20	28.94 (26.27 - 31.76)	2.54 (2.46 - 2.62)	11.01 (8.09 - 14.80)	3.08 (2.57 - 3.69)	28.94 (26.27 - 31.76)	2.54 (2.46 - 2.62)	11.01 (8.09 - 14.80)	3.08 (2.57 - 3.69)
0.50	1.28 (1.24 - 1.33)	0.36 (0.33 - 0.39)	1.09 (1.04 - 1.13)	0.93 (0.85 - 1.02)	1.00	0.36 (0.33 - 0.39)	1.09 (1.04 - 1.13)	0.93 (0.85 - 1.02)
1.00	1.20 (1.16 - 1.25)	0.33 (0.30 - 0.35)	1.01 (0.96 - 1.05)	0.54 (0.51 - 0.58)	2.00	1.20 (1.16 - 1.25)	1.01 (0.96 - 1.05)	0.54 (0.51 - 0.58)
2.00	1.02 (0.98 - 1.06)	0.12 (0.11 - 0.14)	0.82 (0.78 - 0.86)	0.49 (0.45 - 0.52)	5.00	1.02 (0.98 - 1.06)	0.82 (0.78 - 0.86)	0.49 (0.45 - 0.52)
5.00	0.50 (0.47 - 0.53)	0.01 (0.00 - 0.01)	0.30 (0.28 - 0.33)	0.34 (0.32 - 0.37)	10.00	1.02 (0.98 - 1.06)	0.82 (0.78 - 0.86)	0.49 (0.45 - 0.52)
10.00	0.50 (0.47 - 0.53)	0.01 (0.00 - 0.01)	0.30 (0.28 - 0.33)	0.34 (0.32 - 0.37)	20.00	1.02 (0.98 - 1.06)	0.82 (0.78 - 0.86)	0.49 (0.45 - 0.52)
20.00	0.50 (0.47 - 0.53)	0.01 (0.00 - 0.01)	0.30 (0.28 - 0.33)	0.34 (0.32 - 0.37)	50.00	0.50 (0.47 - 0.53)	0.01 (0.00 - 0.01)	0.34 (0.32 - 0.37)

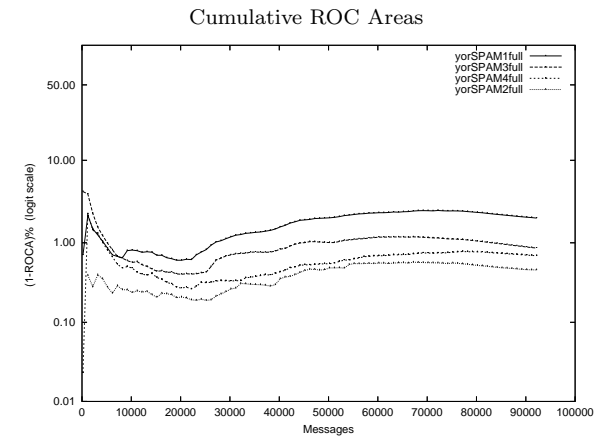
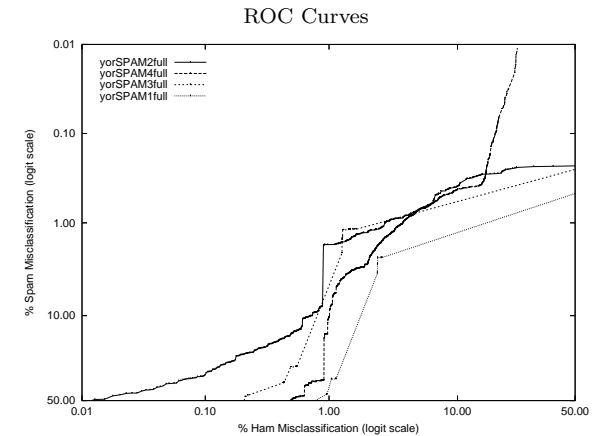


Learning Curves

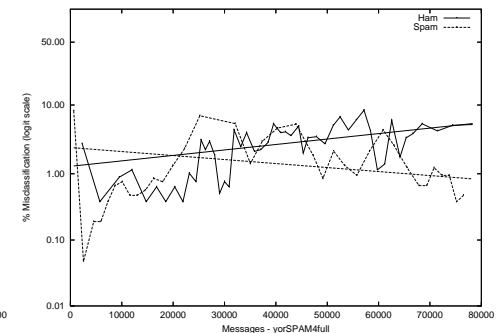
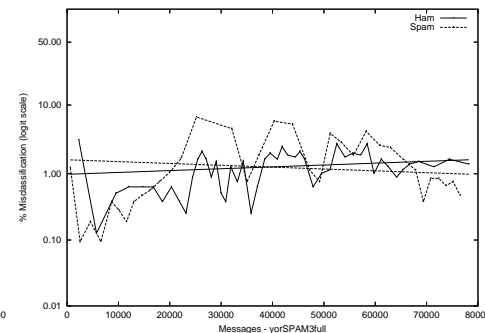
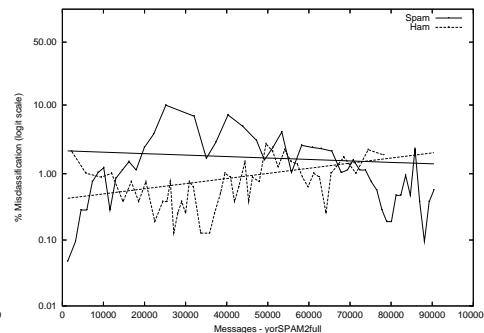
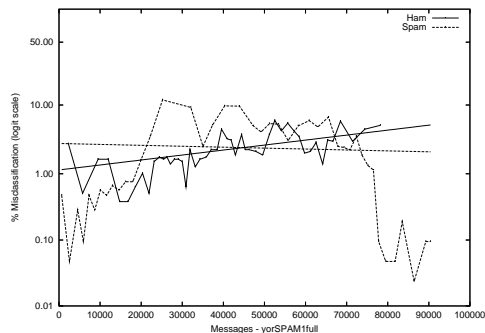


yor - York U. on Public trec05p-1/full Corpus

Gold Std. Filter Classification	yorSPAM1full		yorSPAM2full		yorSPAM3full		yorSPAM4full	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
	38436 963	1281 51509	39035 364	920 51870	38891 508	635 52155	38221 1178	716 52074
Ham Misc%	2.44 (2.29-2.60)		0.92 (0.83-1.02)		1.29 (1.18-1.41)		2.99 (2.82-3.16)	
Spam Misc%	2.43 (2.30-2.56)		1.74 (1.63-1.86)		1.20 (1.11-1.30)		1.36 (1.26-1.46)	
Logit av Misc%	2.44 (2.34-2.54)		1.27 (1.20-1.35)		1.25 (1.18-1.32)		2.02 (1.93-2.11)	
(1-ROCA)%	2.032 (1.939 - 2.129)		0.457 (0.409 - 0.511)		0.861 (0.812 - 0.912)		0.688 (0.636 - 0.745)	
Ham Misc%	Spam Misc%							
0.01	95.61 (87.62 - 98.53)	53.81 (48.00 - 59.52)	82.77 (79.16 - 85.87)	95.75 (89.42 - 98.36)				
0.02	94.77 (93.46 - 95.82)	46.97 (41.07 - 52.95)	80.71 (76.34 - 84.43)	94.66 (93.23 - 95.80)				
0.05	91.80 (90.07 - 93.26)	39.35 (34.84 - 44.05)	71.23 (65.81 - 76.10)	90.36 (85.61 - 93.65)				
0.10	87.24 (83.86 - 90.00)	34.21 (29.93 - 38.77)	62.13 (58.79 - 65.35)	84.92 (81.76 - 87.61)				
0.20	81.04 (79.07 - 82.86)	23.36 (21.00 - 25.91)	52.89 (46.64 - 59.05)	77.23 (73.54 - 80.54)				
0.50	62.71 (57.21 - 67.90)	15.04 (13.51 - 16.71)	29.54 (24.88 - 34.68)	49.67 (42.56 - 56.80)				
1.00	41.54 (34.65 - 48.78)	1.74 (1.06 - 2.84)	13.09 (9.85 - 17.20)	10.42 (4.41 - 22.67)				
2.00	16.38 (12.51 - 21.16)	1.28 (1.18 - 1.40)	1.17 (1.09 - 1.26)	2.89 (2.51 - 3.32)				
5.00	2.37 (2.25 - 2.50)	0.74 (0.67 - 0.82)	1.13 (1.05 - 1.23)	0.74 (0.67 - 0.82)				
10.00	2.24 (2.12 - 2.37)	0.39 (0.34 - 0.46)	1.06 (0.98 - 1.16)	0.43 (0.38 - 0.48)				
20.00	1.96 (1.84 - 2.08)	0.28 (0.24 - 0.33)	0.89 (0.82 - 0.98)	0.06 (0.04 - 0.09)				
50.00	1.09 (1.00 - 1.18)	0.23 (0.19 - 0.28)	0.46 (0.40 - 0.52)	0.00 (0.00 - 0.01)				
Spam Misc%	Ham Misc%							
0.01	93.63 (92.62 - 94.51)	99.76 (nan - nan)	91.58 (89.60 - 93.21)	36.58 (19.37 - 58.07)				
0.02	93.13 (92.30 - 93.88)	99.69 (99.57 - 99.78)	89.85 (88.45 - 91.10)	24.56 (21.30 - 28.14)				
0.05	89.95 (87.96 - 91.64)	99.66 (99.60 - 99.71)	85.24 (82.10 - 87.91)	20.84 (19.25 - 22.52)				
0.10	87.45 (86.08 - 88.71)	99.64 (99.58 - 99.70)	80.51 (78.06 - 82.75)	18.50 (17.54 - 19.50)				
0.20	82.98 (81.06 - 84.74)	99.64 (94.15 - 99.98)	69.44 (66.64 - 72.10)	16.43 (15.83 - 17.04)				
0.50	71.37 (69.13 - 73.51)	6.82 (6.13 - 7.58)	46.67 (42.46 - 50.94)	8.15 (6.94 - 9.55)				
1.00	53.35 (50.04 - 56.64)	2.82 (2.48 - 3.20)	14.20 (9.07 - 21.54)	3.87 (3.58 - 4.18)				
2.00	18.49 (14.56 - 23.20)	0.90 (0.81 - 1.00)	1.27 (1.17 - 1.38)	2.33 (2.18 - 2.49)				
5.00	2.38 (2.22 - 2.56)	0.89 (0.80 - 0.99)	1.18 (1.08 - 1.28)	1.17 (1.07 - 1.27)				
10.00	2.20 (2.04 - 2.37)	0.74 (0.65 - 0.84)	1.07 (0.98 - 1.16)	1.01 (0.92 - 1.10)				
20.00	1.87 (1.73 - 2.01)	0.31 (0.27 - 0.37)	0.79 (0.72 - 0.87)	0.92 (0.84 - 1.01)				
50.00	0.81 (0.73 - 0.90)	0.01 (0.00 - 0.04)	0.21 (0.17 - 0.25)	0.49 (0.43 - 0.56)				

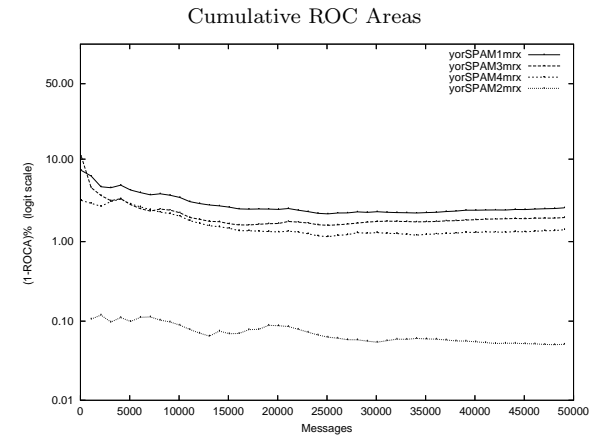
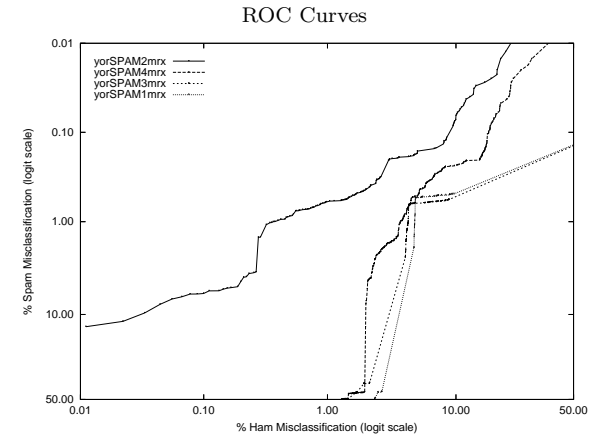


Learning Curves

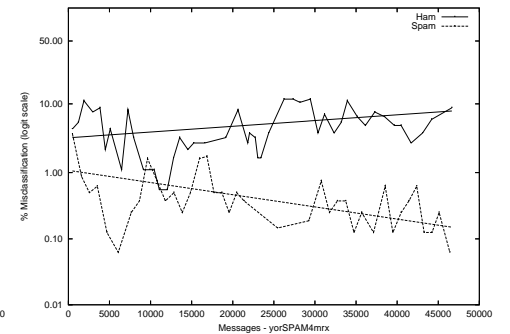
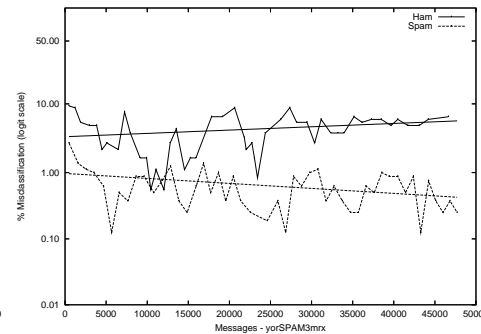
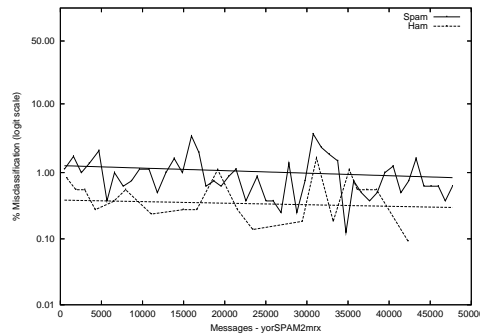
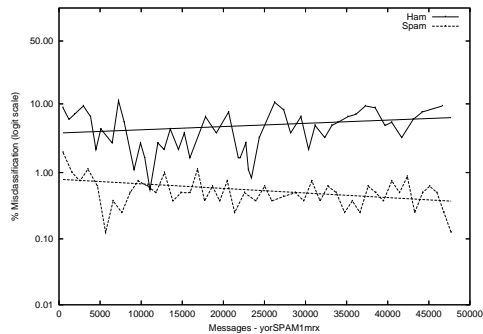


yor - York U. on Private Mr X. Corpus

Gold Std. Filter Classification	yorSPAM1mrx		yorSPAM2mrx		yorSPAM3mrx		yorSPAM4mrx	
	ham	spam	ham	spam	ham	spam	ham	spam
Ham Misc%	8590	219	9007	413	8639	260	8568	179
Spam Misc%	448	39829	31	39635	399	39788	470	39869
Logit av Misc%	4.96 (4.52-5.42)		0.34 (0.23-0.49)		4.41 (4.00-4.86)		5.20 (4.75-5.68)	
(1-ROCA)%	0.55 (0.48-0.62)		1.03 (0.93-1.14)		0.65 (0.57-0.73)		0.45 (0.38-0.52)	
	1.67 (1.54-1.80)		0.60 (0.50-0.71)		1.71 (1.58-1.85)		1.55 (1.42-1.68)	
Ham Misc%	2.632 (2.407 - 2.878)		0.051 (0.039 - 0.067)		1.993 (1.800 - 2.205)		1.407 (1.191 - 1.662)	
					Spam Misc%			
0.01	98.93 (98.43 - 99.28)	42.83 (13.06 - 78.88)	98.48 (95.71 - 99.47)	98.91 (98.42 - 99.25)				
0.02	98.77 (97.71 - 99.34)	13.22 (3.47 - 39.28)	95.56 (88.24 - 98.41)	98.84 (97.74 - 99.41)				
0.05	97.66 (95.95 - 98.66)	7.88 (4.75 - 12.79)	93.91 (91.08 - 95.88)	96.79 (93.35 - 98.48)				
0.10	95.76 (93.49 - 97.26)	6.08 (4.99 - 7.38)	92.07 (88.17 - 94.76)	96.18 (94.41 - 97.41)				
0.20	92.95 (90.87 - 94.59)	4.55 (2.32 - 8.71)	86.26 (81.63 - 89.87)	91.45 (85.80 - 94.98)				
0.50	88.44 (85.73 - 90.69)	0.86 (0.70 - 1.07)	75.20 (70.99 - 78.98)	84.13 (76.03 - 89.86)				
1.00	77.95 (71.89 - 83.01)	0.60 (0.51 - 0.70)	60.19 (54.93 - 65.23)	70.64 (51.29 - 84.61)				
2.00	57.14 (51.36 - 62.74)	0.45 (0.37 - 0.55)	39.91 (35.44 - 44.55)	32.01 (3.17 - 87.13)				
5.00	0.54 (0.09 - 3.32)	0.17 (0.13 - 0.23)	0.62 (0.55 - 0.70)	0.52 (0.43 - 0.63)				
10.00	0.49 (0.43 - 0.56)	0.07 (0.04 - 0.12)	0.56 (0.49 - 0.65)	0.23 (0.19 - 0.28)				
20.00	0.43 (0.37 - 0.49)	0.02 (0.01 - 0.04)	0.48 (0.41 - 0.55)	0.06 (0.04 - 0.09)				
50.00	0.25 (0.21 - 0.30)	0.00 (0.00 - 0.01)	0.30 (0.25 - 0.36)	0.00 (0.00 - 0.02)				
					Ham Misc%			
0.01	97.22 (89.01 - 99.34)	23.68 (17.62 - 31.03)	91.71 (77.66 - 97.24)	33.92 (22.17 - 48.07)				
0.02	95.68 (89.58 - 98.28)	19.35 (13.91 - 26.27)	88.43 (83.24 - 92.16)	26.36 (18.45 - 36.15)				
0.05	85.61 (80.47 - 89.57)	11.36 (8.93 - 14.35)	83.08 (80.00 - 85.77)	20.24 (17.38 - 23.43)				
0.10	75.86 (69.38 - 81.33)	9.22 (8.12 - 10.44)	76.48 (71.80 - 80.59)	16.98 (15.83 - 18.21)				
0.20	59.44 (52.39 - 66.12)	3.11 (1.94 - 4.95)	62.64 (56.38 - 68.49)	14.82 (10.67 - 20.20)				
0.50	9.02 (4.07 - 18.78)	1.68 (1.27 - 2.23)	16.88 (8.75 - 30.10)	5.22 (4.54 - 6.01)				
1.00	4.90 (4.49 - 5.35)	0.37 (0.24 - 0.55)	4.36 (3.97 - 4.78)	3.78 (3.37 - 4.24)				
2.00	4.82 (4.41 - 5.28)	0.28 (0.19 - 0.41)	4.18 (3.80 - 4.60)	2.73 (2.39 - 3.12)				
5.00	4.71 (4.29 - 5.18)	0.20 (0.13 - 0.31)	3.99 (3.63 - 4.40)	2.08 (1.76 - 2.45)				
10.00	4.39 (3.98 - 4.84)	0.03 (0.01 - 0.11)	3.70 (3.33 - 4.10)	2.02 (1.71 - 2.39)				
20.00	3.87 (3.50 - 4.29)	0.01 (0.00 - 0.04)	3.12 (2.78 - 3.50)	2.01 (1.70 - 2.38)				
50.00	2.38 (2.11 - 2.68)	0.01 (0.01 - 0.01)	1.55 (1.31 - 1.83)	1.28 (1.00 - 1.64)				

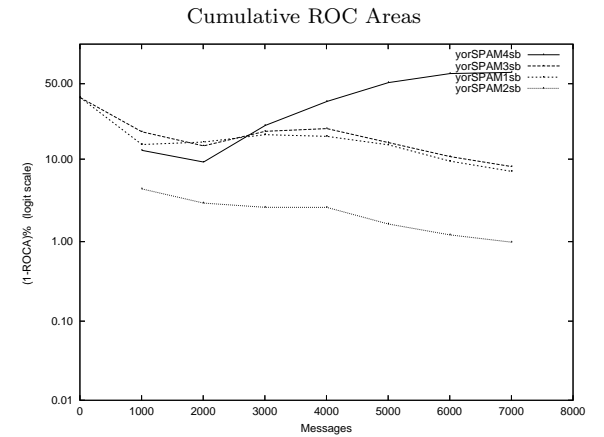
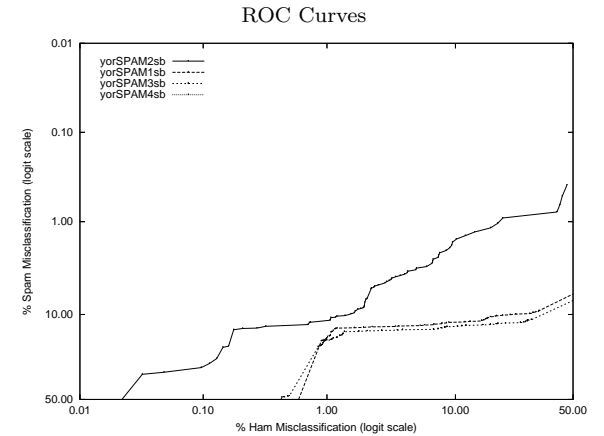


Learning Curves

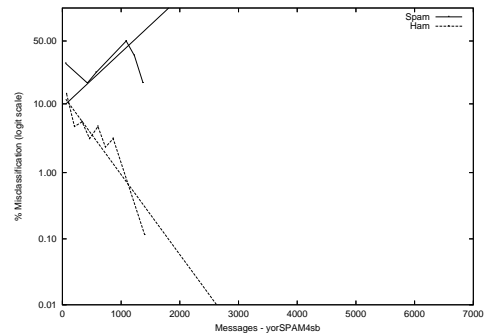
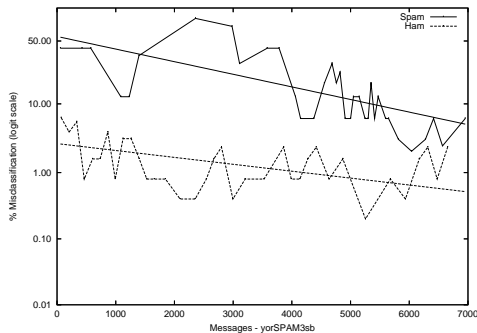
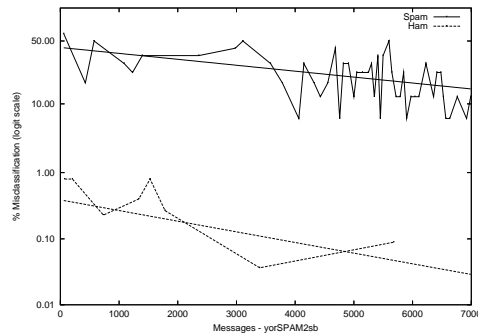
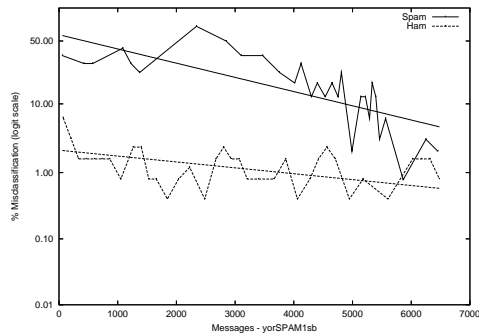


yor - York U. on Private S.B. Corpus

Gold Std. Filter Classification	yorSPAM1sb			yorSPAM2sb			yorSPAM3sb			yorSPAM4sb		
	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	spam	ham spam	ham spam	
Ham Misc%	1.19 (0.93-1.49)			0.14 (0.07-0.27)			1.36 (1.09-1.68)			0.77 (0.57-1.02)		
Spam Misc%	13.81 (11.45-16.44)			23.64 (20.68-26.80)			15.32 (12.85-18.07)			91.35 (89.15-93.24)		
Logit av Misc%	4.20 (3.63-4.87)			2.07 (1.49-2.88)			4.76 (4.15-5.47)			22.26 (19.16-25.71)		
(1-ROCA)%	7.237 (5.997 - 8.708)			0.983 (0.657 - 1.468)			8.234 (6.871 - 9.839)			58.165 (56.047 - 60.254)		
Ham Misc%	Spam Misc%											
0.01	92.00 (85.40 - 95.77)			71.17 (47.00 - 87.29)			82.21 (77.35 - 86.21)			99.74 (nan - nan)		
0.02	86.71 (75.30 - 93.32)			62.99 (30.91 - 86.62)			82.21 (75.57 - 87.34)			99.10 (nan - nan)		
0.05	78.45 (67.11 - 86.66)			33.12 (17.94 - 52.86)			77.79 (67.82 - 85.34)			98.45 (95.58 - 99.47)		
0.10	77.16 (71.37 - 82.08)			30.52 (18.43 - 46.07)			70.13 (58.45 - 79.67)			98.06 (95.81 - 99.12)		
0.20	74.06 (66.86 - 80.17)			14.16 (8.10 - 23.57)			61.95 (54.61 - 68.78)			96.77 (93.84 - 98.34)		
0.50	53.29 (44.34 - 62.04)			13.12 (10.82 - 15.82)			47.79 (34.26 - 61.65)			93.68 (90.30 - 95.93)		
1.00	17.03 (10.77 - 25.87)			11.95 (9.55 - 14.84)			17.79 (13.59 - 22.94)			90.71 (88.12 - 92.78)		
2.00	13.42 (11.13 - 16.09)			8.18 (5.48 - 12.05)			14.68 (12.48 - 17.18)			90.45 (87.83 - 92.55)		
5.00	13.03 (10.78 - 15.67)			3.38 (2.25 - 5.04)			14.16 (11.99 - 16.64)			89.81 (87.15 - 91.96)		
10.00	12.00 (9.87 - 14.51)			1.69 (0.99 - 2.85)			13.25 (11.13 - 15.69)			88.90 (86.11 - 91.19)		
20.00	10.32 (8.36 - 12.68)			1.04 (0.52 - 2.08)			12.34 (10.22 - 14.82)			87.35 (84.53 - 89.72)		
50.00	6.97 (5.48 - 8.81)			0.39 (0.11 - 1.33)			8.05 (6.34 - 10.17)			63.61 (59.67 - 67.38)		
Spam Misc%	Ham Misc%											
0.01	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.02	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.05	100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)			100.00 (nan - nan)		
0.10	95.78 (93.94 - 97.08)			96.89 (62.00 - 99.83)			96.60 (94.57 - 97.89)			99.89 (nan - nan)		
0.20	94.43 (92.10 - 96.11)			62.80 (15.52 - 93.94)			94.71 (91.37 - 96.80)			99.87 (nan - nan)		
0.50	94.09 (91.14 - 96.11)			47.16 (24.44 - 71.13)			93.59 (90.59 - 95.67)			99.78 (99.44 - 99.91)		
1.00	90.85 (85.18 - 94.49)			21.15 (6.06 - 52.75)			90.97 (87.91 - 93.32)			99.52 (98.99 - 99.77)		
2.00	84.51 (77.45 - 89.66)			9.06 (5.35 - 14.93)			86.26 (78.67 - 91.44)			98.73 (97.80 - 99.27)		
5.00	64.08 (53.97 - 73.08)			2.57 (1.77 - 3.71)			70.17 (61.27 - 77.77)			96.34 (95.07 - 97.29)		
10.00	25.53 (9.84 - 51.87)			1.56 (0.82 - 2.94)			41.31 (27.42 - 56.73)			93.52 (91.79 - 94.90)		
20.00	0.88 (0.69 - 1.13)			0.18 (0.09 - 0.34)			0.88 (0.65 - 1.20)			85.43 (83.16 - 87.43)		
50.00	0.61 (0.45 - 0.83)			0.03 (0.01 - 0.12)			0.43 (0.29 - 0.64)			61.76 (59.00 - 64.44)		

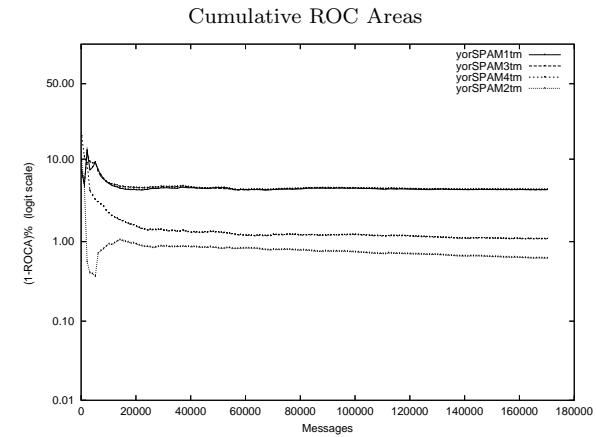
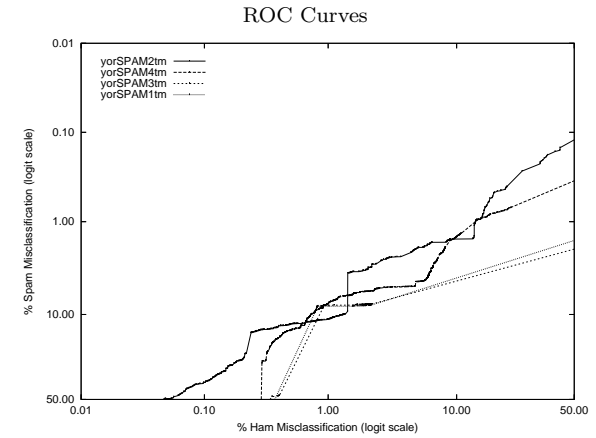


Learning Curves

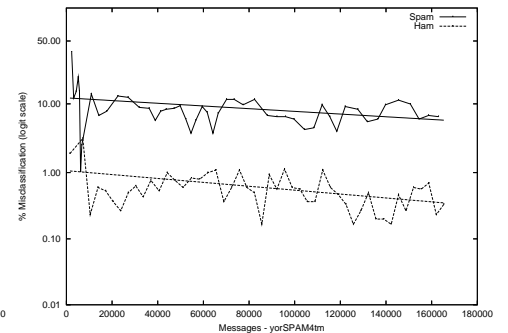
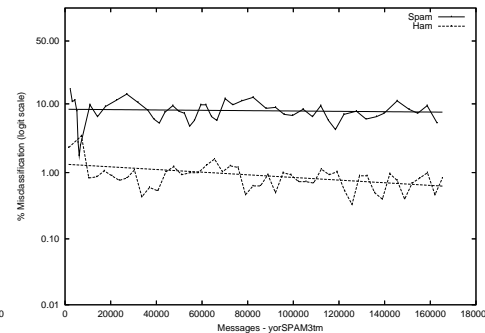
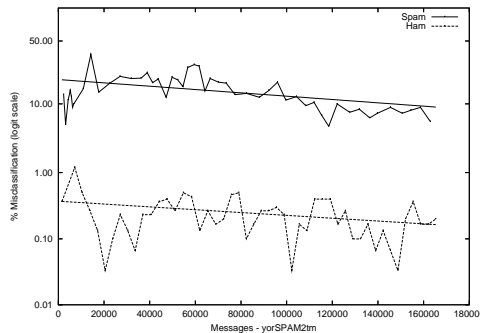
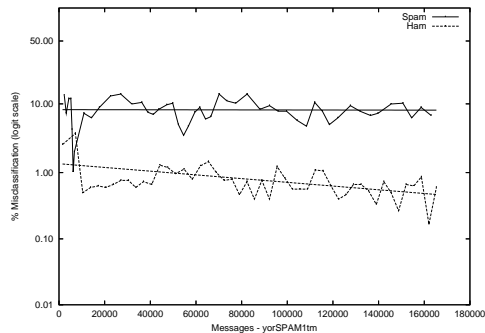


yor - York U. on Private T.M. Corpus

Gold Std. Filter Classification	yorSPAM1tm		yorSPAM2tm		yorSPAM3tm		yorSPAM4tm	
	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam	ham spam
Ham Misc%	0.82 (0.77-0.86)		0.25 (0.22-0.28)		0.92 (0.87-0.97)		0.63 (0.59-0.67)	
Spam Misc%	8.28 (7.89-8.67)		14.90 (14.40-15.41)		8.11 (7.73-8.50)		9.04 (8.64-9.45)	
Logit av Misc%	2.65 (2.56-2.75)		2.05 (1.94-2.16)		2.78 (2.69-2.89)		2.45 (2.35-2.54)	
(1-ROCA)%	4.400 (4.184 - 4.626)		0.619 (0.567 - 0.677)		4.366 (4.132 - 4.612)		1.081 (1.004 - 1.162)	
Ham Misc%	Spam Misc%							
0.01	93.69 (92.35 - 94.81)	67.46 (61.45 - 72.95)	93.30 (91.17 - 94.95)	95.16 (93.82 - 96.23)				
0.02	92.28 (90.61 - 93.67)	59.13 (55.27 - 62.88)	89.90 (87.93 - 91.57)	93.74 (92.48 - 94.81)				
0.05	86.64 (84.30 - 88.68)	49.40 (46.03 - 52.78)	84.37 (82.42 - 86.15)	87.24 (83.30 - 90.36)				
0.10	78.76 (75.09 - 82.01)	39.19 (36.96 - 41.47)	78.42 (75.83 - 80.80)	78.66 (75.49 - 81.52)				
0.20	67.01 (63.59 - 70.25)	26.62 (23.62 - 29.87)	65.73 (63.76 - 67.65)	58.23 (55.88 - 60.54)				
0.50	38.68 (35.03 - 42.46)	12.51 (11.98 - 13.06)	40.35 (37.95 - 42.80)	14.38 (13.72 - 15.07)				
1.00	8.25 (7.85 - 8.67)	11.17 (10.73 - 11.62)	8.04 (7.64 - 8.46)	7.81 (7.30 - 8.36)				
2.00	8.13 (7.73 - 8.55)	3.41 (3.18 - 3.66)	7.93 (7.54 - 8.33)	5.86 (5.54 - 6.20)				
5.00	7.80 (7.43 - 8.20)	2.00 (1.81 - 2.22)	7.56 (7.17 - 7.96)	4.57 (4.27 - 4.88)				
10.00	7.36 (7.01 - 7.73)	1.55 (1.39 - 1.74)	7.11 (6.74 - 7.50)	1.49 (1.33 - 1.66)				
20.00	6.51 (6.17 - 6.86)	0.46 (0.36 - 0.57)	6.34 (5.97 - 6.73)	0.80 (0.69 - 0.93)				
50.00	4.14 (3.87 - 4.42)	0.14 (0.09 - 0.21)	4.10 (3.82 - 4.39)	0.49 (0.41 - 0.59)				
Spam Misc%	Ham Misc%							
0.01	99.80 (nan - nan)	98.36 (95.19 - 99.45)	99.73 (99.44 - 99.87)	99.86 (98.59 - 99.99)				
0.02	99.71 (98.21 - 99.95)	94.76 (85.75 - 98.20)	99.57 (98.91 - 99.83)	98.41 (90.27 - 99.76)				
0.05	98.34 (96.17 - 99.29)	85.42 (73.39 - 92.57)	98.55 (96.78 - 99.35)	95.34 (90.86 - 97.68)				
0.10	97.45 (96.52 - 98.14)	68.67 (50.17 - 82.68)	97.72 (97.17 - 98.16)	90.73 (84.87 - 94.47)				
0.20	96.00 (95.41 - 96.51)	36.06 (26.77 - 46.52)	96.81 (96.38 - 97.20)	78.99 (71.65 - 84.83)				
0.50	93.56 (92.53 - 94.46)	18.07 (15.51 - 20.94)	94.47 (93.73 - 95.13)	48.91 (40.03 - 57.85)				
1.00	86.94 (85.04 - 88.64)	13.69 (12.64 - 14.83)	89.18 (87.27 - 90.84)	13.71 (11.47 - 16.31)				
2.00	75.14 (72.77 - 77.37)	5.04 (4.24 - 5.98)	76.34 (73.69 - 78.81)	8.24 (7.90 - 8.60)				
5.00	38.57 (34.84 - 42.45)	1.43 (1.37 - 1.50)	38.04 (34.14 - 42.09)	4.88 (3.75 - 6.34)				
10.00	0.79 (0.75 - 0.84)	1.33 (1.19 - 1.47)	0.88 (0.84 - 0.93)	0.73 (0.68 - 0.79)				
20.00	0.70 (0.66 - 0.74)	0.23 (0.21 - 0.25)	0.76 (0.72 - 0.80)	0.35 (0.32 - 0.37)				
50.00	0.38 (0.35 - 0.42)	0.05 (0.04 - 0.06)	0.34 (0.31 - 0.36)	0.29 (0.27 - 0.32)				



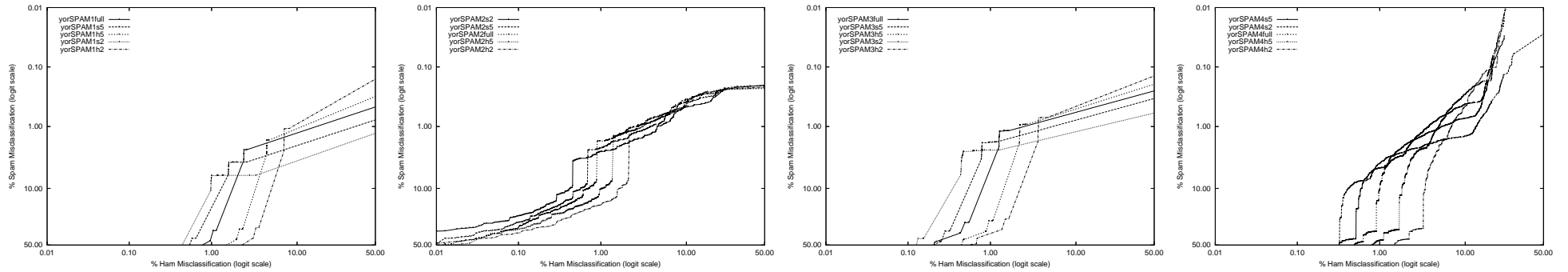
Learning Curves



yor - York U. - Public subset Corpora

	yorSPAM1full	yorSPAM1h2	yorSPAM1h5	yorSPAM1s2	yorSPAM1s5
Ham Misc%	2.44 (2.29-2.60)	7.22 (6.71-7.75)	4.55 (4.27-4.86)	1.00 (0.91-1.11)	1.62 (1.49-1.75)
Spam Misc%	2.43 (2.30-2.56)	1.08 (1.00-1.18)	1.70 (1.59-1.81)	6.36 (5.95-6.79)	3.89 (3.66-4.13)
Logit av Misc%	2.44 (2.34-2.54)	2.84 (2.69-3.00)	2.79 (2.67-2.92)	2.56 (2.41-2.71)	2.51 (2.39-2.64)
(1-ROCA)%	2.032 (1.939 - 2.129)	3.292 (3.031 - 3.574)	2.564 (2.411 - 2.726)	3.234 (3.011 - 3.474)	2.369 (2.223 - 2.524)
	yorSPAM2full	yorSPAM2h2	yorSPAM2h5	yorSPAM2s2	yorSPAM2s5
Ham Misc%	0.92 (0.83-1.02)	2.26 (1.97-2.57)	1.45 (1.29-1.63)	0.48 (0.41-0.55)	0.72 (0.64-0.81)
Spam Misc%	1.74 (1.63-1.86)	1.17 (1.08-1.27)	1.44 (1.34-1.54)	3.60 (3.29-3.94)	2.43 (2.24-2.62)
Logit av Misc%	1.27 (1.20-1.35)	1.63 (1.51-1.76)	1.44 (1.35-1.55)	1.32 (1.22-1.44)	1.32 (1.23-1.42)
(1-ROCA)%	0.457 (0.409 - 0.511)	0.669 (0.602 - 0.745)	0.530 (0.484 - 0.580)	0.420 (0.349 - 0.504)	0.426 (0.368 - 0.493)
	yorSPAM3full	yorSPAM3h2	yorSPAM3h5	yorSPAM3s2	yorSPAM3s5
Ham Misc%	1.29 (1.18-1.41)	3.75 (3.38-4.15)	2.26 (2.06-2.48)	0.47 (0.41-0.54)	0.80 (0.71-0.89)
Spam Misc%	1.20 (1.11-1.30)	0.72 (0.65-0.79)	0.95 (0.87-1.04)	2.60 (2.33-2.88)	1.86 (1.70-2.04)
Logit av Misc%	1.25 (1.18-1.32)	1.65 (1.54-1.77)	1.47 (1.38-1.56)	1.11 (1.02-1.21)	1.22 (1.14-1.31)
(1-ROCA)%	0.861 (0.812 - 0.912)	1.382 (1.264 - 1.512)	1.082 (0.993 - 1.180)	1.176 (1.054 - 1.311)	1.025 (0.946 - 1.110)
	yorSPAM4full	yorSPAM4h2	yorSPAM4h5	yorSPAM4s2	yorSPAM4s5
Ham Misc%	2.99 (2.82-3.16)	9.98 (9.39-10.59)	5.66 (5.34-6.00)	0.96 (0.87-1.06)	1.74 (1.61-1.87)
Spam Misc%	1.36 (1.26-1.46)	0.48 (0.42-0.54)	0.77 (0.70-0.85)	3.87 (3.55-4.21)	2.32 (2.14-2.51)
Logit av Misc%	2.02 (1.93-2.11)	2.26 (2.11-2.42)	2.11 (2.00-2.24)	1.94 (1.82-2.07)	2.01 (1.90-2.12)
(1-ROCA)%	0.688 (0.636 - 0.745)	1.975 (1.771 - 2.201)	1.117 (1.003 - 1.244)	0.586 (0.525 - 0.654)	0.537 (0.491 - 0.587)

ROC Curves



Cumulative ROC Area

