KDD2001, Binding to Thrombin

	Precision	Recall	F-Score
SVM Performance on full data set, c = 0.1:	0.2941	0.0333	0.0599
Reduced set 1(10%):			
SVM Performance on the reduced data set:	0.1429	0.0067	0.0127
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.1818	0.12	0.1446
SVM Spy-EM:	0	0	0
Quality of Entity Set Expansion:			
Bayesian Sets:	0.23	0.6053	0.3333
Spy-EM:	0.27	0.7105	0.3913
Reduced set 2 (33%):			
SVM Performance on the reduced data set:	0.1579	0.02	0.0355
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.1667	0.1133	0.1349
SVM Spy-EM:	0.1183	0.0733	0.0905
Quality of Entity Set Expansion:			
Bayesian Sets:	0.1684	0.5517	0.2581
Spy-EM:	0.2	0.6552	0.3065
Reduced set 3 (80%):			
SVM Performance on the reduced data set:	0.1905	0.0267	0.0468
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.1649	0.1067	0.1296
SVM Spy-EM:	0.1348	0.08	0.1004
Quality of Entity Set Expansion:			
Bayesian Sets:	0.0854	0.7778	0.1538
Spy-EM:	0.0854	0.7778	0.1538

KDD2001, Binding to Thrombin

	Precision	Recall	F-Score
SVM Performance on full data set, c = 0.9:	0.2941	0.0333	0.0599
Reduced set 1(10%):			
SVM Performance on the reduced data set:	0.1111	0.0067	0.0126
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.2143	0.18	0.1957
SVM Spy-EM:	0.2143	0.18	0.1957
Quality of Entity Set Expansion:			
Bayesian Sets:	0.2637	0.6316	0.3721
Spy-EM:	0.2857	0.6842	0.4031
Reduced set 2 (33%):			
SVM Performance on the reduced data set:	0.25	0.0267	0.0482
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.2366	1	0.3827
SVM Spy-EM:	0.2366	1	0.3827
Quality of Entity Set Expansion:			
Bayesian Sets:	0.1939	0.6552	0.2992
Spy-EM:	0.2245	0.7586	0.3465
Reduced set 3 (80%):			
SVM Performance on the reduced data set:	0.3571	0.0333	0.061
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.1707	0.0933	0.1207
SVM Spy-EM:	0	0	0
Quality of Entity Set Expansion:			
Bayesian Sets:	0.0602	0.5556	0.1087
Spy-EM:	0.0723	0.6667	0.1304

Reuters 21578, text classification:

	Precision	Recall	F-Score
SVM Performance on full data set, c = 0.1:	0.9601	0.8935	0.9256

Same thrends for c = 0.5, for 0.9 both fail to achieve much augmentation.

Reduced set 1(10%):

SVM Performance on the reduced data set:	0.9525	0.679	0.7928
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.4762	0.8887	0.6201
SVM Spy-EM:	0.7707	0.9919	0.8674
Quality of Entity Set Expansion:			
Bayesian Sets:	0.6636	1	0.7978
Spy-EM:	0.6626	0.9985	0.7965

Reduced set 2 (33%):

SVM Performance on the reduced data set:	0.9488	0.8065	0.8718
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.5673	0.8839	0.691
SVM Spy-EM:	0.8904	0.9565	0.9222
Quality of Entity Set Expansion:			
Bayesian Sets:	0.5866	1	0.7394
Spy-EM:	0.5854	0.9979	0.7379

Reduced set 3 (80%):

SVM Performance on the reduced data set:	0.9582	0.8871	0.9213
SVM Performance on the augmented data sets:			
SVMAugmented Bayes:	0.7766	0.8581	0.8153
SVM Spy-EM:	0.9265	0.8742	0.8996
Quality of Entity Set Expansion:			
Bayesian Sets:	0.2816	1	0.4395
Spy-EM:	0.2816	1	0.4395

20 Newsgroups, c = 0.5.

Augmenting with Bayesian sets:

1st redux: 0.9375 0.0085 0.0169
aug: 0.4366 0.9778 0.6037
2nd redux: 0.9756 0.0228 0.0445
aug 0.4353 0.9727 0.6014
3rd redux 0.5461 0.9704 0.6989
aug 0.3742 0.9949 0.5439
full train: 0.4984 0.9858 0.6621

Augmenting with S-EM algo:

1st redux: 0.9375 0.0085 0.0169
aug 0.3460 0.9966 0.5136
2nd redux: 0.9756 0.0228 0.0445
aug 0.2870 1.0000 0.4461
3rd redux: 0.5461 0.9704 0.6989
aug 0.3994 0.9983 0.5705
full train: 0.4984 0.9858 0.6621

Precision, recall and fscore of the entity set expansion used to get aug:

bayes1: 0.7593 0.8463 0.8004
sem1: 0.7775 0.8666 0.8196

bayes2: 0.6777 0.8931 0.7706
sem2: 0.6761 0.8910 0.7688

bayes3: 0.3263 0.9456 0.4851
sem3: 0.3251 0.9421 0.4833

Bayes extracts exactly as many new positives as with SEM, using SEM's number as the cutoff.

20 Newsgroups, c = 0.1, same P:N ratio in all training data:

Augmenting with bayes:

redux1: 0.9547 0.2641 0.4137
aug: 0.9860 0.0803 0.1484
redux2: 1.0000 0.0028 0.0057
aug: 0.7047 0.9289 0.8014
redux3: 0.7161 0.9317 0.8098
aug: 0.6208 0.9596 0.7539
full train: 0.7542 0.9118 0.8256

Augmenting with SEM:

redux1: 0.9547 0.2641 0.4137
aug: 0.9809 0.0876 0.1609
redux2: 1.0000 0.0028 0.0057
aug: 0.7101 0.9311 0.8057
redux3: 0.7161 0.9317 0.8098
aug: 0.6071 0.9664 0.7457
full train: 0.7542 0.9118 0.8256

Precision, recall and fscore of the entity set expansion used to get aug:

bayes: 0.7558 0.8631 0.8059
sem: 0.7681 0.8771 0.8190

bayes: 0.6543 0.9031 0.7588
sem: 0.6619 0.9136 0.7676

bayes: 0.2737 0.9632 0.4262
sem: 0.2672 0.9404 0.4161

20 Newsgroups, c = 0.9 (to get less overfitting, but a terrible classifier, for redux1/2):

Augmenting with Bayes:

redux1: 1.0000 0.0023 0.0045
aug: 0.9762 0.0233 0.0456
redux2: 0.3301 0.9829 0.4943
aug: 0.2847 1.0000 0.4432
redux3: 0.2848 0.9892 0.4422
aug: 0.3181 0.9954 0.4822
full train: 0.2860 0.9960 0.4444

Augmenting with SEM:

redux1: 1.0000 0.0023 0.0045
aug: 0.8642 0.0398 0.0762
redux2: 0.3301 0.9829 0.4943
aug: 0.2904 0.9989 0.4500
redux3: 0.2848 0.9892 0.4422
aug: 0.2883 0.9994 0.4475
full train: 0.2860 0.9960 0.4444

Precision, recall and fscore of the entity set expansion used to get aug:

bayes: 0.7609 0.8526 0.8041
sem: 0.7772 0.8709 0.8214

bayes: 0.6585 0.9062 0.7628
sem: 0.6608 0.9094 0.7654

bayes: 0.2730 0.9737 0.4264
sem: 0.2646 0.9439 0.4134

When kkt = 0.9, the classifier is so bad that extra data doesn't help much at all.
That's why, apart from augmenting redux1, it makes only marginal improvement in the rest.