

1. Part 1 (Perceptron)

For convenience, all results are included in output.zip file, though size of the folder is big.

To evaluate the performance of Perceptron for various given datasets, we tried with various iterations and learning rates.

For iteration we select the values:

50,100,150,200,250,300,350,400,450,500,550,600,650,700,750,800,850,900,950, and 1000.

And for iterations, the selected values are: 0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0.

The results are provided in the tables.

Comparison with Naive Bayes and Logistic Regression:

Enron1

Naive Bayes

For this dataset, Naive Bayes receives 94.2982 % accuracy with following confusion matrix.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	290	17
	Spam	9	140

However, we received highest accuracy 94.51% using smoothing value 2 instead of 1. And in this case, only one more spam correctly detected.

Logistic Regression

We did the same experiments for LR with same iterations and learning rates.

The best accuracy is received with iteration 850 after eliminating stop words and the accuracy is 94.95%. In this case, 141 spams out of 149 and 292 hams out of 307 are detected correctly.

Both NB and LR outperforms Perceptron (highest accuracy is 92.10%). Though Perceptron converges quickly (as we increase iteration, the accuracy remain same), it does not show better performance over these two classifiers.

Enron4

Naive Bayes

NB (removing stop words) outperforms Perceptron where accuracy for NB is 96.31 % (correctly detected spams 384 out of 391 and hams 139 out of 152) and accuracy for Perceptron is 93.92% (correctly detected spams 373 out of 391 and hams 134 out of 152).

Logistic Regression Surprisingly LR performs best among three of these classifiers. And accuracy is 97.97% where correctly detecting hams 390 out of 391 and spams 142 out of 152. And we get it for learning rate 0.01 and iteration number is 450.

hw2

Naive Bayes In hw2 dataset, NB is also dominating over Perceptron. Accuracy of NB is 94.76% (detects 125 spams correctly out of 130 and 328 hams correctly out of 348). On the other hand, accuracy of Perceptron is 90.79% (detects 109 spams out of 130 and 325 hams out of 348).

Logistic Regression

Still LR outperforms Perceptron by receiving accuracy 93.72% and detects 117 spams correctly out of 130 and 331 hams correctly out of 348.

Overall, in all datasets NB and LR outperform Perceptron. As Perceptron has less parameters than LR and both NB and LR can deal with non-linear decision surface, it may be cannot exceed these two classifiers. For dataset Enron4, LR outperforms NB. The reason can be LR is less biased than NB.

2. Part 2 (Neural Network)

Hidden Layer

For hidden layer, the parameter setting is given below:

Options: -L (learning rate) 0.01 -M (Momentum) 0.01 -N (Iteration) 500 -V 0 -S 0 -E 20 -H (Hidden layer) 1

We try, from 1 hidden layer to 5 hidden layers.

To make report shorter, we present only the best result for each dataset.

Enron1: Best accuracy is received for hidden layer 2 and it is 91.886 %. Confusion matrix is given below as well.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	300	7
	Spam	30	119

As we increase hidden layer, it does not increase monotonically. Higher hidden layer does not bring any benifit.

Enron4: For this dataset we use N (Number of iteration) = 600. Best accuracy is received for hidden layer 4 and it is 92.2652 %. Confusion matrix is given below as well.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	110	42
	Spam	0	391

Higher hidden layer helps this dataset to get better result. Probably, the decision boundary is very non-linear.

hw2: Best accuracy is received for hidden layer 1 (and 5) and it is 94.1423 %. Confusion matrix is given below as well.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	328	20
	Spam	8	122

For hidden layer 5, it detects more 10 hams correctly and 10 spam incorrectly.

Hidden Unit:

For this we keep fix hidden layer 2. And the options are:

Options: -L 0.01 -M 0.01 -N 500 -V 0 -S 0 -E 20 -H 2 (Hidden Layer), 1 (Hidden Unit)

We run the experiments for hidden unit: 1, 2, 3, 4, 5, 10, 15, and 20. For Enron4 dataset number of iteration is 600.

Enron1: Best accuracy is received when number of hidden units are 10 (layer 2) and it is 90.7895 %. Confusion matrix is given below as well.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	302	5
	Spam	37	112

Still, we do not see any monotonic behavior for increasing hidden units.

Enron4: For this dataset, for hidden units performance are almost same 72.0074 %. Noticeably, almost all hidden can not detect ham at all. Hidden unit 20 can identify only one ham. And almost for hidden units, we receive 100% accuracy for spam. That means hypothesis works like a random classifier.

hw2: hw2 dataset receives best accuracy for hidden unit 4 and it is 93.5146 %. Confusion matrix is given below as well.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	342	6
	Spam	43	87

Learning Rate:

We tried for following settings:

Options: -L 1e-4 -M 0.01 -N 500 -V 0 -S 0 -E 20 -H 1

We tested for learning rates: 0.0, 0.01, 0.05, 0.1, 0.3, 0.5, 0.8, and 1.0. Where 0.0 and 1.0 are used for references.

Enron1:

For Enron1, we receive best accuracy (89.4737%) for 0.01. Which is not promising. Probably other parameters are not well tuned as well. In other learning cases, the performance is even poor. Classifier can detect all as ham.

Enron4:

For this data set, we do not see any effect on the performance. In all cases, classifiers detect all test examples as spam. Probably this dataset has large number of parameters and require more iterations to understand the effect of learning rate or other factors. The confusion matrix looks like as follows.

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	0	152
	Spam	0	391

hw2: with learning rate 0.01, we receive the highest accuracy 94.1423 %. The confusion matrix is as follows:

		Prediction outcome	
		Ham'	Spam'
actual value	Ham	328	20
	Spam	8	122

For other cases, the classifiers detect all examples as Spam.

Overall, for small value of learning rate, performance is better. Probably, we need to change learning rate more finely rather than big value.

Momentum:

With momentum, we try with following option.

Options: -L 0.01 -M given Value -N 500 -V 0 -S 0 -E 20 -H 1 And we tried with same values as learning rate.

for Enron4 dataset, we try with 600 iteration.

Enron1: The highest accuracy in this dataset for momentum change is 89.693%, received for 0.0 and 0.05. And for other cases, accuracy difference is not very high.

Enron4:

For this dataset, for all cases, the accuracy is same 72.0074 % and all test examples are detected as spam.

hw2

With 0.05 and 0.1, we receive highest accuracy 94.3515 % but the confusion matrix is different for each cases. For 0.1, 9 more hams are detected correctly and 9 less spams are detected incorrectly.

Iteration:

To understand the effect of iteration, we set the option as following way.

Options: -L 0.3 -M 0.2 -N given value -V 0 -S 0 -E 20 -H 1

The given values are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, and 500. As Enron4 dataset has large parameters, we tried with more parameters for this dataset.

Enron1:

We get a linear relation for iteration numbers and accuracy. Upto 100 iteration we do not see change that means the classifier detects all examples as hams. And for 500, we receive highest accuracy, 89.4737 %.

Enron4:

For this dataset, we tried upto 1000 iteration. However, we do not find linear relation like earlier dataset. The highest accuracy is received for 92.081 % for iteration number 600. For higher iterations, the model probably over learn.

hw2: Like Enron1, we get a linear relation with accuracy and iteration numbers. And receive highest, accuracy for iteration 500 and which is 94.1423 %. Probably, if we increase iteration more, the model will over-learn.

3. Part 3 (Collaborative Filtering)

Mean Absolute Error: 0.7632759782819423

Root Mean Squared Error: 0.9300823189029057

Table 1: Perceptron with various iterations (enron1 – with stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
100	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
150	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
200	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
250	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
300	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
350	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
400	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
450	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
500	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
550	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
600	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
650	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
700	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
750	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
800	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
850	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
900	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
950	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969
1000	Spam (149)	128	0.8590
	Ham (307)	281	0.9153
	Total (456)	409	0.8969

Table 2: Perceptron with various iterations (enron1 – without stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Spam (149)	139	0.9328
	Ham (307)	279	0.9087
	Total (456)	418	0.9166
100	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
150	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
200	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
250	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
300	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
350	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
400	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
450	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
500	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
550	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
600	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
650	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
700	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
750	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
800	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
850	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
900	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
950	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
1000	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122

Table 3: Perceptron with various learning rates (enron1 – with stop word)

learning rate	Classification	Correct	Detection	Accuracy
0.0	Spam (149)	149	1.0	
	Ham (307)	0	0.0	
	Total (456)	149	0.3267	
0.01	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.02	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.03	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	409	0.8969	
0.04	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.05	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	406	0.8903	
0.06	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	406	0.8903	
0.07	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.08	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.09	Spam (149)	129	0.8657	
	Ham (307)	280	0.9120	
	Total (456)	409	0.8969	
0.1	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	406	0.8903	
0.2	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	406	0.8903	
0.3	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.4	Spam (149)	128	0.8590	
	Ham (307)	127	0.8523	
	Total (456)	406	0.8903	
0.5	Spam (149)	131	0.8791	
	Ham (307)	280	0.9120	
	Total (456)	411	0.9013	
0.6	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	409	0.8969	
0.7	Spam (149)	128	0.8590	
	Ham (307)	282	0.9185	
	Total (456)	410	0.8991	
0.8	Spam (149)	127	0.8523	
	Ham (307)	279	0.9087	
	Total (456)	406	0.8903	
0.9	Spam (149)	128	0.8590	
	Ham (307)	281	0.9153	
	Total (456)	410	0.8991	
1.0	Spam (149)	131	0.8791	
	Ham (307)	280	0.9120	
	Total (456)	411	0.9013	

Table 4: Perceptron with various learning rates (enron1 – without stop word)

Iteration	Classification	Correct Detection	Accuracy
0.0	Spam (149)	149	1.0
	Ham (307)	0	0.0
	Total (456)	418	0.9166
0.01	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
0.02	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
0.03	Spam (149)	138	0.9261
	Ham (307)	281	0.9153
	Total (456)	419	0.9188
0.04	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
0.05	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
0.06	Spam (149)	138	0.9261
	Ham (307)	281	0.9153
	Total (456)	419	0.9188
0.07	Spam (149)	139	0.9328
	Ham (307)	275	0.8957
	Total (456)	414	0.9078
0.08	Spam (149)	134	0.8993
	Ham (307)	282	0.9185
	Total (456)	416	0.9122
0.09	Spam (149)	133	0.8926
	Ham (307)	278	0.9055
	Total (456)	411	0.9013
0.1	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
0.2	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
0.3	Spam (149)	144	0.9664
	Ham (307)	275	0.8957
	Total (456)	419	0.9188
0.4	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
0.5	Spam (149)	136	0.9127
	Ham (307)	282	0.9185
	Total (456)	418	0.9166
0.6	Spam (149)	144	0.9664
	Ham (307)	275	0.8957
	Total (456)	419	0.9188
0.7	Spam (149)	141	0.9463
	Ham (307)	280	0.9120
	Total (456)	421	0.9232
0.8	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
0.9	Spam (149)	139	0.9328
	Ham (307)	281	0.9153
	Total (456)	420	0.9210
1.0	Spam (149)	136	0.9127
	Ham (307)	282	0.9185
	Total (456)	418	0.9166

Table 5: Perceptron with various iterations (enron4 – with stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
100	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
150	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
200	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
250	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
300	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
350	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
400	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
450	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
500	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
550	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
600	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
650	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
700	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
750	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
800	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
850	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
900	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
950	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
1000	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392

Table 6: Perceptron with various iterations (enron4 – without stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
100	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
150	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
200	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
250	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
300	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
350	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
400	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
450	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
500	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
550	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
600	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
650	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
700	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
750	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
800	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
850	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
900	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
950	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
1000	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392

Table 7: Perceptron with various learning rates (enron4 – with stop word)

learning rate	Classification	Correct Detection	Accuracy
0.0	Ham (152)	0	0.0
	Spam (391)	391	1.0
	Total (543)	391	0.7200
0.01	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.02	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.03	Ham (152)	138	0.9078
	Spam (391)	360	0.9207
	Total (543)	498	0.9171
0.04	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.05	Ham (152)	138	0.9078
	Spam (391)	355	0.9079
	Total (543)	493	0.9079
0.06	Ham (152)	138	0.9078
	Spam (391)	360	0.9207
	Total (543)	498	0.9171
0.07	Ham (152)	137	0.9013
	Spam (391)	359	0.9181
	Total (543)	496	0.9134
0.08	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.09	Ham (152)	139	0.9144
	Spam (391)	369	0.9437
	Total (543)	508	0.9355
0.1	Ham (152)	138	0.9078
	Spam (391)	355	0.9079
	Total (543)	493	0.9079
0.2	Ham (152)	138	0.9078
	Spam (391)	355	0.9079
	Total (543)	493	0.9079
0.3	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.4	Ham (152)	138	0.9078
	Spam (391)	355	0.9079
	Total (543)	493	0.9079
0.5	Ham (152)	139	0.9144
	Spam (391)	369	0.9437
	Total (543)	508	0.9355
0.6	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.7	Ham (152)	139	0.9144
	Spam (391)	367	0.9386
	Total (543)	506	0.9318
0.8	Ham (152)	138	0.9078
	Spam (391)	355	0.9079
	Total (543)	493	0.9079
0.9	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
1.0	Ham (152)	139	0.9144
	Spam (391)	355	0.9079
	Total (543)	494	0.9079

Table 8: Perceptron with various learning rates (enron4 – without stop word)

Iteration	Classification	Correct Detection	Accuracy
0.0	Ham (152)	0	0.0
	Spam (391)	391	1.0
	Total (543)	391	0.7200
0.01	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.02	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.03	Ham (152)	138	0.9078
	Spam (391)	367	0.9386
	Total (543)	505	0.9300
0.04	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.05	Ham (152)	136	0.8947
	Spam (391)	366	0.9360
	Total (543)	502	0.9244
0.06	Ham (152)	138	0.9078
	Spam (391)	367	0.9386
	Total (543)	505	0.9300
0.07	Ham (152)	138	0.9078
	Spam (391)	365	0.9335
	Total (543)	503	0.9263
0.08	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.09	Ham (152)	139	0.9144
	Spam (391)	368	0.9411
	Total (543)	507	0.9337
0.1	Ham (152)	136	0.8947
	Spam (391)	366	0.9360
	Total (543)	502	0.9244
0.2	Ham (152)	136	0.8947
	Spam (391)	366	0.9360
	Total (543)	502	0.9244
0.3	Ham (152)	139	0.9144
	Spam (391)	367	0.9386
	Total (543)	506	0.9318
0.4	Ham (152)	136	0.8947
	Spam (391)	366	0.9360
	Total (543)	502	0.9244
0.5	Ham (152)	139	0.9144
	Spam (391)	368	0.9411
	Total (543)	507	0.9337
0.6	Ham (152)	139	0.9144
	Spam (391)	367	0.9386
	Total (543)	506	0.9318
0.7	Ham (152)	137	0.9013
	Spam (391)	373	0.9539
	Total (543)	510	0.9392
0.8	Ham (152)	136	0.8947
	Spam (391)	366	0.9360
	Total (543)	502	0.9244
0.9	Ham (152)	138	0.9078
	Spam (391)	365	0.9335
	Total (543)	503	0.9263
1.0	Ham (152)	139	0.9144
	Spam (391)	367	0.9386
	Total (543)	506	0.9318

Table 9: Perceptron with various iterations (hw2 – with stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
100	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
150	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
200	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
250	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
300	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
350	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
400	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
450	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
500	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
550	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
600	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
650	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
700	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
750	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
800	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
850	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
900	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
950	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016
1000	Spam (130)	108	0.8307
	Ham (348)	323	0.9281
	Total (478)	431	0.9016

Table 10: Perceptron with various iterations (hw2 – without stop word)

Iteration	Classification	Correct Detection	Accuracy
50	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
100	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
150	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
200	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
250	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
300	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
350	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
400	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
450	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
500	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
550	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
600	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
650	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
700	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
750	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
800	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
850	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
900	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
950	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079
1000	Spam (130)	109	0.8384
	Ham (348)	325	0.9339
	Total (478)	434	0.9079

Table 11: Perceptron with various learning rates (hw2 – with stop word)

learning rate	Classification	Correct	Detection	Accuracy
0.0	Spam (130)	130	1.0	
	Ham (348)	0	0.0	
	Total (478)	130	0.2719	
0.01	Spam (130)	108	0.8307	
	Ham (348)	323	0.9281	
	Total (478)	431	0.9016	
0.02	Spam (130)	108	0.8307	
	Ham (348)	323	0.9281	
	Total (478)	431	0.9016	
0.03	Spam (130)	109	0.9078	
	Ham (348)	320	0.9195	
	Total (478)	429	0.8974	
0.04	Spam (130)	108	0.8307	
	Ham (348)	323	0.9281	
	Total (478)	431	0.9016	
0.05	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.06	Spam (130)	109	0.9078	
	Ham (348)	320	0.9195	
	Total (478)	429	0.8974	
0.07	Spam (130)	112	0.8615	
	Ham (348)	320	0.9195	
	Total (478)	432	0.9037	
0.08	Spam (130)	108	0.8307	
	Ham (348)	323	0.9281	
	Total (478)	431	0.9016	
0.09	Spam (130)	113	0.8692	
	Ham (348)	314	0.9022	
	Total (478)	427	0.8933	
0.1	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.2	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.3	Spam (130)	112	0.8615	
	Ham (348)	320	0.9195	
	Total (478)	432	0.9037	
0.4	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.5	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.6	Spam (130)	112	0.8615	
	Ham (348)	320	0.9195	
	Total (478)	432	0.9037	
0.7	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.8	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	
0.9	Spam (130)	111	0.9013	
	Ham (348)	319	0.9166	
	Total (478)	430	0.8538	
1.0	Spam (130)	110	0.8461	
	Ham (348)	320	0.9195	
	Total (478)	430	0.8995	

Table 12: Perceptron with various learning rates (hw2 – without stop word)

Iteration	Classification	Correct	Detection	Accuracy
0.0	Spam (130)	0		1.0
	Ham (348)	0		0.0
	Total (478)	130		0.2719
0.01	Spam (130)	109		0.8384
	Ham (348)	325		0.9339
	Total (478)	434		0.9079
0.02	Spam (130)	109		0.8384
	Ham (348)	325		0.9339
	Total (478)	434		0.9079
0.03	Spam (130)	110		0.8461
	Ham (348)	320		0.9195
	Total (478)	430		0.8995
0.04	Spam (130)	109		0.8384
	Ham (348)	325		0.9339
	Total (478)	434		0.9079
0.05	Spam (130)	109		0.8384
	Ham (348)	323		0.9281
	Total (478)	432		0.9037
0.06	Spam (130)	110		0.8461
	Ham (348)	320		0.9195
	Total (478)	430		0.8995
0.07	Spam (130)	110		0.8461
	Ham (348)	320		0.9195
	Total (478)	430		0.8995
0.08	Spam (130)	109		0.8384
	Ham (348)	325		0.9339
	Total (478)	434		0.9079
0.09	Spam (130)	109		0.8384
	Ham (348)	321		0.9224
	Total (478)	430		0.9037
0.1	Spam (130)	109		0.8384
	Ham (348)	323		0.9281
	Total (478)	432		0.9037
0.2	Spam (130)	109		0.8384
	Ham (348)	323		0.9281
	Total (478)	432		0.9037
0.3	Spam (130)	108		0.8307
	Ham (348)	317		0.9109
	Total (478)	425		0.8891
0.4	Spam (130)	109		0.8384
	Ham (348)	323		0.9281
	Total (478)	432		0.9037
0.5	Spam (130)	110		0.8461
	Ham (348)	320		0.9195
	Total (478)	430		0.8995
0.6	Spam (130)	108		0.8307
	Ham (348)	317		0.9109
	Total (478)	425		0.8891
0.7	Spam (130)	109		0.8384
	Ham (348)	320		0.9195
	Total (478)	429		0.8974
0.8	Spam (130)	109		0.8384
	Ham (348)	323		0.9281
	Total (478)	432		0.9037
0.9	Spam (130)	107		0.8230
	Ham (348)	323		0.9281
	Total (478)	430		0.8995
1.0	Spam (130)	110		0.8461
	Ham (348)	320		0.9195
	Total (478)	430		0.8995