

Homework 2 Report

To see how to run the two projects, read the readme file presented.

Naïve Bayes Classifier

For the three given data sets(hw2_train/test, enron1_train/test, enron4_train/test), an accuracy displayed in the following picture was obtained.

Naïve Bayes Classifier															
Test Data Results		Ham Correct Predictions		Ham Incorrect Predictions		Spam Correct Predictions		Spam Incorrect Predictions		Total Number		Accurate		Accuracy	
DS1		338		10		121		2		471		459		0.974	
DS2		301		6		131		0		438		432		0.986301	
DS 3		130		22		382		20		554		512		0.924188	

The program is a straightforward Python implementation of the Multinomial Naïve Bayes classifier from the given link in the HW description.

Logistic Regression

Since Logistic Regression Algorithm takes a high amount of time to learn, it was tested on different parameters for low number of iterations and the results are presented in the Excel sheet available in the project folder. Accuracy on the test data that was obtained is presented below.

Test Set									
DS1		1	0.01	0.01	198	140	121	1	0.6934
DS2		1	0.01	0.01	209	107	129	2	0.75615
DS3		1	0.01	0.01	14	119	394	0	0.774194

Perceptron Algorithm

For Perceptron algorithm, it was found that for approximately 750 iterations, the accuracy on the validation set is maximized. Results for all the 3 datasets are available in the Excel sheet in the Java Project folder and one can observe that the accuracy on the validation set tends to maximize around 750 iterations.

Perceptron:		
DS 1		
For Validation Set:		
No Of Iterations	Accuracy	
1	0.79	
50	0.84	
100	0.85	
250	0.86	
500	0.86	
750	0.87	
1000	0.85	
DS2		
For Validation Set:		
No Of Iterations	Accuracy	
1	0.8222	
50	0.814814	
100	0.837	
250	0.8444	
500	0.866667	
750	0.84444	
1000	0.82962	

DS3			
For Validation Set			
No Of Iterations	Accuracy		
1	0.7826		
50	0.9192		
100	0.8944		
250	0.92		
500	0.88		
750	0.937		
1000	0.826		
For Test Set			
DS1			
No Of Iterations	Accuracy of Ham	Accuracy of Spam	
750	0.91954023	0.9	
DS2			
750	0.912052117	0.88590604	
DS3			
750	0.881578947	0.938618926	

Validation was performed by training the Perceptron only on Training data and testing was performed by training the perceptron on both training and validation data.