Report of Naive Bayes Classifier and Logistic Regression Classifier

Name : Karthika Karunakaran
Net Id : kxk152430

Naive Bayes

	Spam Words in Vocabulary bag	Ham Words in Vocabulary bag
With Stop Words	6217	5839
Without Stop Words	5822	5422

Accuracy of Spam/Ham classification for test data = 77.61507

Accuracy of Spam/Ham classification for test data without stop words = 85.35565

Logistic Regression

Accuracy of the test data with stop words= 88.66489832007073

With Stop Words:

Learning Rate	Regularization constant	No of Iterations	Accuracy
0.018	0.05	500	75.00442086648984
0.018	0.8	500	87.79840848806367
0.018	0.1	500	78.41954022988506
0.018	0.9	500	89.8187444739169
0.018	0.025	500	81.59372236958444

Accuracy of the test data without stop words= 90.10610079575596

Without Stop Words:

Learning Rate	Regularization constant	No of Iterations	Accuracy
0.018	0.05	500	86.11626878868259
0.018	0.8	500	88.42396109637488
0.018	0.1	500	86.78824049513705
0.018	0.9	500	89.99593280282937
0.018	0.025	500	79.57338638373122

Keeping the Learning rate constant and at an iteration limit of 500 where convergence happens, accuracy is calculated for different values of Regularization constant(lambda) and results are found above. The Regularization constant of 0.9 gives greater accuracy in this case.