Machine Learning 6375.002

Assignment 2

Problem:

Spam classifier is implemented using

- Naïve Bayes Algorithm
- Logistic Regression

Naïve Bayes Implementation:

A conditional probability based Naïve bayes is implemented. The implementation is done with stop words removal and without stop word removal.

Results:

```
****************************************
Spam classifier using Naive Bayes (With Stop Words)
****************************************
Accuracy for Ham
97.7011 %
Accuracy for Spam
85.3846 %
****************************************
Spam classifier using Naive Bayes (Without stopWords)
****************************************
Accuracy for Hpam
96.5517 %
Accuracy for Spam
90.7692 %
```

Conclusion:

After removing stop words, the accuracy of the spam classifier was significantly improved.

Logistic Regression Implementation:

Results:

Learning Rate	Regularization Factor	Iterations	Spam Accuracy	Ham Accuracy
0.025	0.9	100	63%	98.85%
0.0025	1	50	53.38%	97.70%
0.25	0.1	150	86.15%	95.68%

Conclusion:

The spam classifier's accuracy is improved when the number of iterations are kept above 100, provided the regularization parameter, learning rate are maintained between 0.1 to 0.5.