Machine Learning
Homework – 2
Classification problem – Naive Bayes and Logistic Regression

Vivek Anand Sampath
vxs135130

Naive Bayes:

WORD SET USED FOR TRAINING	Accuracy (%)
Including all words	96.23
Filtering all stop words	95.815

Filtering out stop words does not affect Naive Bayes classification much.

Logistic Regression:

Learning rate (eta)	Regression constant (lambda)	Iterations for Gradient ascent	Filter stop words flag	Accuracy (%)
0.0001	5	10	True	82.63
0.0003	5	10	True	84.72
0.001	5	10	True	89.53
0.003	5	10	True	90.37
0.01	5	10	True	91.42
0.01	15	10	True	89.53
0.01	10	10	True	61.50
0.01	2	10	True	92.050
0.01	1	10	True	89.330
0.01	0.5	10	True	84.93
0.01	2	10	False	83.89

The filtering out of stop words works for Logistic Regression because it reduces the number of unnecessary iterations done for unnecessary features which doesn't affect the criteria and including stop words also reduces the time required for convergence.