

# MACHIINE LEARNING

## Assignment 3

### Naïve Bayes:

Number of Spam files in training set : 123

Number of ham files in training set : 340

Number of Spam files in testing set : 130

Number of ham files in testing set : 348

No of Train words - spam : 24470

No of unique Train words - spam : 6289

No of Train words - ham : 74414

No of unique Train words - ham : 5904

Confusion Matrix

	spam	ham
Classified spam	129	1
Classified ham	31	317

Accuracy : 93.30543933054393 %

### Logistic Regression:

#### Changing Learning Rate:

Number of iterations : 5000

Learning Rate : 0.1

Lambda : 0.5

Training Accuracy: 73.4341252699784

Testing Accuracy: 72.80334728033473

Number of iterations : 5000

Learning Rate : 0.1

Lambda : 0.2

Training Accuracy: 73.4341252699784

Testing Accuracy: 72.80334728033473

Number of iterations : 5000

Learning Rate : 0.1  
Lambda : 0.1  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.01  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.001  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473

### **Changing Lambda:**

Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.5  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.2  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.1  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.01  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473  
Number of iterations : 5000  
Learning Rate : 0.1  
Lambda : 0.001  
Training Accuracy: 73.4341252699784  
Testing Accuracy: 72.80334728033473

Neither changing learning rate nor lambda changed the train accuracy and test accuracy.

## **Naïve Bayes - No Stop Words:**

Confusion Matrix

	spam	ham
Classified spam	129	1
Classified ham	70	278

Accuracy - No Stop Words : 85.14644351464436 %

## **Logistic Regression - No Stop Words:**

After removing stop words logistic regression gave same accuracy for both train and test accuracy for different values of learning rate and lambda.

Training Accuracy: 73.4341252699784

Testing Accuracy: 72.80334728033473