**TUNING THE HYPERPARAMETER FOR LOGISTIC REGRESSION**

The hyperparameter for LOGISTIC REGRESSION i.e lambda(L) was tuned (0.0001, 0.001, 0.005, 0.01, 0.05) by taking different values of learning rate (N) ranging from 0.001 to 10 and increasing by a factor of 10 each time.

The difference between the accuracy on the 70% split training set and the accuracy on the 30% split validation set was observed for various values of L and different epochs.

The number of epochs were kept as 100, 200, 300, 500, 1000.

**1.Which data representation and algorithm combination yields the best performance (mea-**

**sured in terms of the accuracy, precision, recall and F1 score) and why?**

Naïve bayes yields the best performance with the bag of words model because it has the frequency of the words which gives essentially more input to the model.

**2. Does Multinomial Naive Bayes perform better (again performance is measured in terms**

**of the accuracy, precision, recall and F1 score) than LR and SGDClassifier on the Bag**

**of words representation? Explain your yes/no answer.**

Yes it has more it has more accuracy because of frequency of words.

**3. Does Discrete Naive Bayes perform better (again performance is measured in terms of**

**the accuracy, precision, recall and F1 score) than LR and SGDClassifier on the Bernoulli**

**representation? Explain your yes/no answer.**

Discrete Naïve bayes does not perform better as there is only the indication of presence of data.