**Report:**

**To compile the program:** javac -Xlint NaiveBayes.java

To run the program: java NaiveBayes

**To compile the program:** javac LogReg.java

To run the program: java LogReg

**Stop Words:**

By removing the stop words the accuracy should generally increase as the stop words are the most common words that can occur in any kind of documents and by including their probability does not contribute to the correct accuracy of the prediction. But it has decreased by removing the stop words.

**Accuracy:**

**Logistic Regression:**

Accuracy on training set is 85%

Accuracy on test set is 90%

**Using Lambda value for L2 Regularization:**

* Regularization is the technique used to reduce overfitting. “Regularization” is done to modify the optimization problem to prefer small weights.
* Regularization is done in following steps

1. Run the program for range of different ʎs using only the training set.
2. For each of the ʎ fit in step 2, check how well the resulting weights fit the test data.
3. Output the weights that perform best on test data.

* To prevent the magnitude of weight values from becoming large, the idea of regularization is to penalize weight values by adding those weight values to the calculation of the error term.
* If weight values are included in the total error term that’s being minimized, then smaller weight values will generate smaller error values.
* L2 weight regularization penalizes weight values by adding the sum of their squared values to the error term. Large weights can lead to overfitting, which leads to poor prediction accuracy.
* Regularization limits the magnitude of model weights by adding a penalty for weights to the model error function. L2 regularization uses the sum of the squared values of the weights.
* ʎ trades off how much we care about the model fitting well versus how we care about weight being small.
* Different values of ʎ used is 0.3, 0.27,0.56,0.4

**Iterations:**

* The number of Iterations needs to be done until the data converge. The Number of Iterations needs to be considered so that the data doesn’t start diverging. More number of Iteration leads to diverging of data.

**Naive Bayes:**

* Accuracy on training set is 88%
* Accuracy on test set is 80%

In Naive Bayes If a particular feature/word does not appear in a particular class, then its conditional probability is equal to 0. If we use the product of probabilities the product becomes 0, while if we use the sum of their logarithms the log(0) is undefined. To avoid this, we will use add-one or Laplace smoothing by adding 1 to each count.