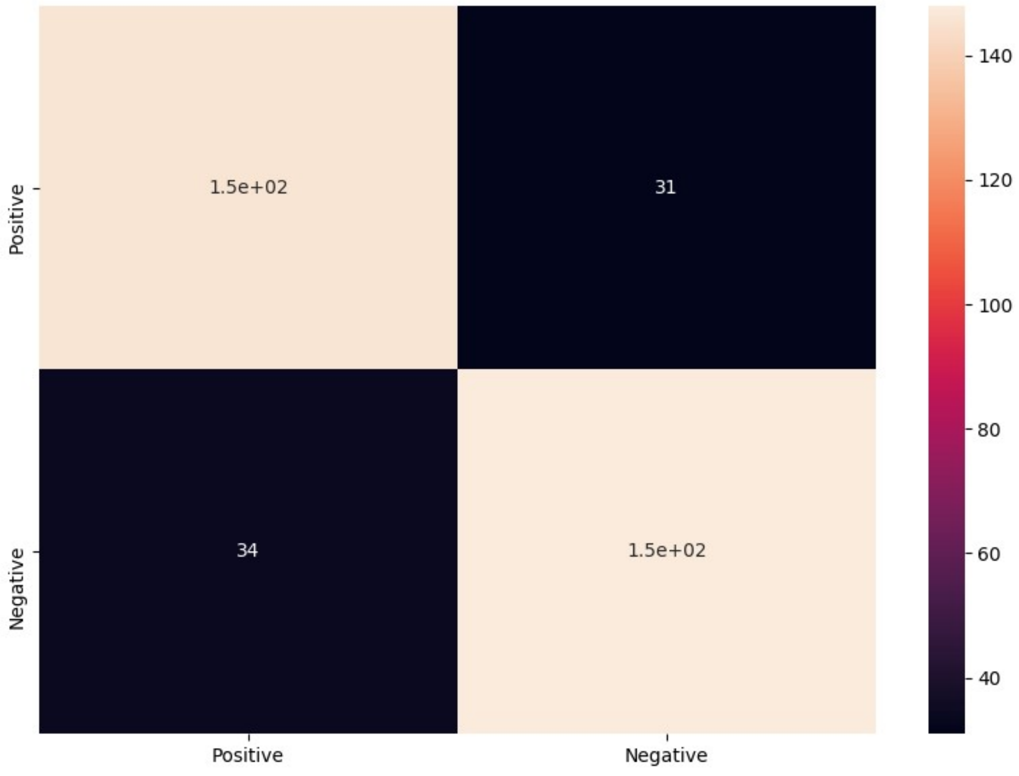
COL 774: Assignment 2(Part A)

a. Accuracy Over Train Data : 86.2645625
Accuracy Over Test Data : 81.8941504178273

b. Random Prediction accuracy = 0.4986072423398329
Majority Prediction accuracy = 0.5069637883008357

Algorithm gives 1.6 times better accuracy than random/majority prediction accuracy .

c. Confusion Matrix



Confusion Matrix : [[146 31]
[34 148]]

Negative(i.e y = '0') has highest value among the diagonal entries .

Element C[i,j] is the number of times article i is mis-classified as article j .

d . Acurracy Over Test Data : 81.05849582172702

After Stemming and removal of stop words there is very less change in the accuracy .

e . Accuracy Over Test Data : 82.17270194986072 (Feature : Unigram + Bigram)

Accuracy Over Test Data : 77% (Feature : Only Bigram)

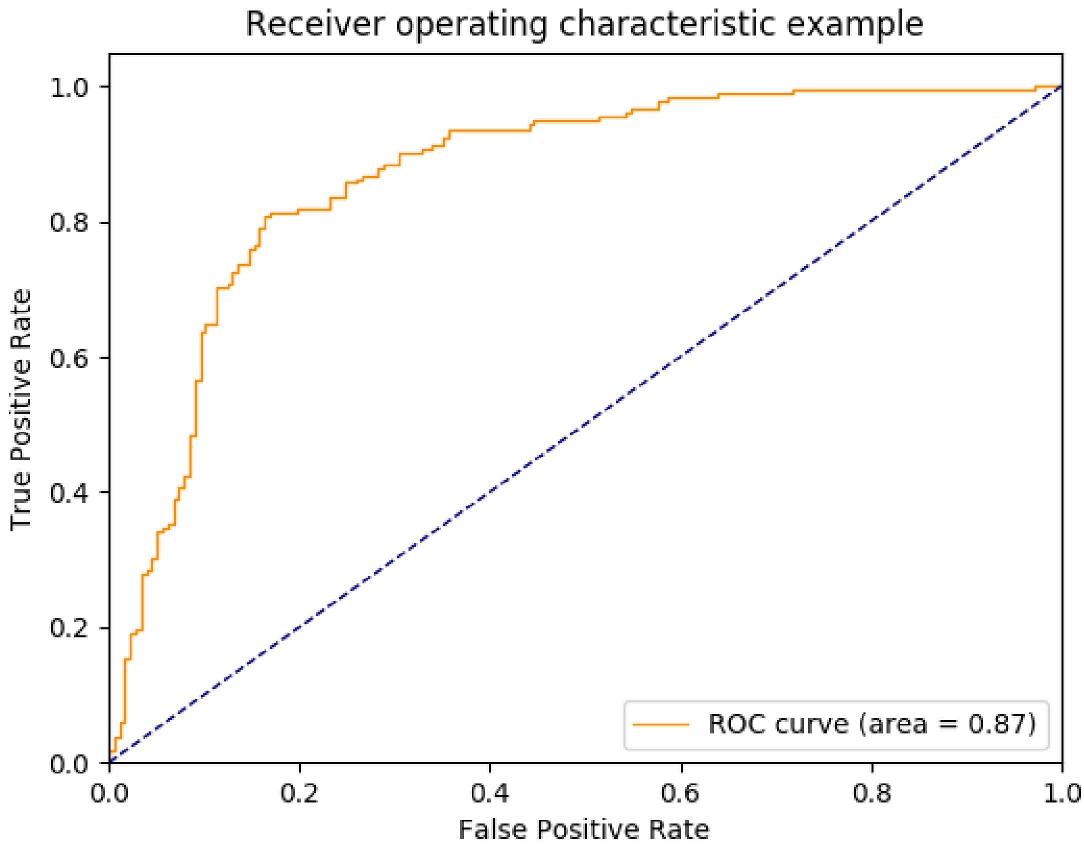
After trying many feature engineering techniques, there is no significant improvement over baseline Multinomial Naive Bayes model. Even with stopword removal and stemming, there is no improvement over the baseline model.

f. 1. Accuracy = 56%

2. Accuracy = 75%

selecting smaller set of features improves time taken to train the model . Moreover selecting top 10 % percentile of the features based on the scoring function helps in increasing the accuracy .

g.



ROC curve for y = 1 . class 4 is the true class.

From the curve we notice that decreasing the prbability threshold to predict value increases the TPR drastically while the FPR does not increase as much till a point . Till this point our model have been said to be overly biased towards class 0.

After a point , decreasing the probability threshold has the opposite effect. TPR doesn't increase as much , but FPR does. After this point , our models is overly biased towards class 1

COL 774: Assignment 2 (Part B)

1. a)

b = [-0.09955545]

Total Number of SVM: 89

Val Accuracy : 99.6%

Test Accuracy : 99.9%

b.)

b = [-635.92785287]

Total Number of SVM: 1243

Val Accuracy : 99%

Test Accuracy : 99.4%

c.) SVC(kernel="rbf", gamma=0.05)

test accuracy : 99.8%

val accuracy : 99.2 %

Number of SVM : 1236

SVC(kernel='linear')

test accuracy : 99.9%

Number of SVM : 89

val accuracy : 99.6 %

Time required for training using the Scikit SVM library is much less when compared with my implementation of binary classification SVM.

Accuracy and number of SVM's and w,b obtained from part a,b are similar to those of part c .

2.)

a.)Accuracy over test data : 80%

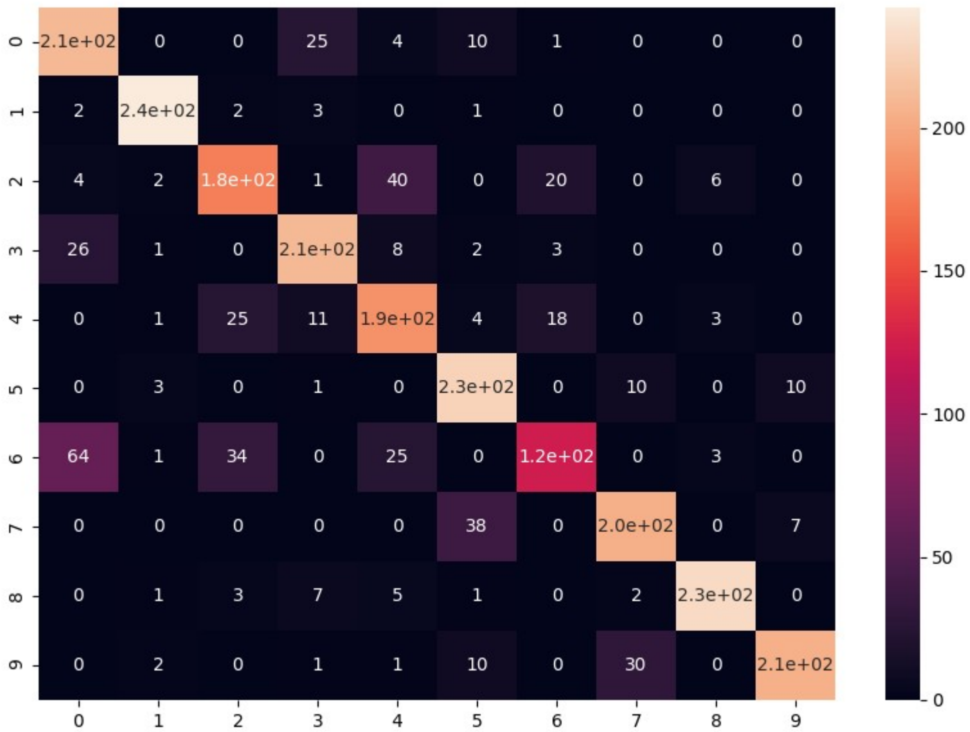
Accuracy over validation data : 79.32%

b) Accuracy over test data : 83%

Accuracy over validation data : 82.92%

Time required for training using the Scikit SVM library is much less when compared with my implementation of one vs one Multiclass SVM. Moreover it gives better accuracy

c.)



Confusion Matrix For Val Data (using my own implementation of one vs one SVM multiclassification)

[[210 ,0 , 0 , 25 , 4 , 10 , 1 , 0 , 0 , 0],

[2, 242 , 2 , 3 , 0 , 1 , 0 , 0 , 0 , 0],

[4 , 2 , 177 , 1 , 40 , 0 , 20 , 0 , 6 , 0],

[26 ,1, 0, 210, 8, 2, 3, 0 , 0 , 0],

[0 , 1 ,25, 11 ,188 , 4 , 18 , 0 , 3 , 0],

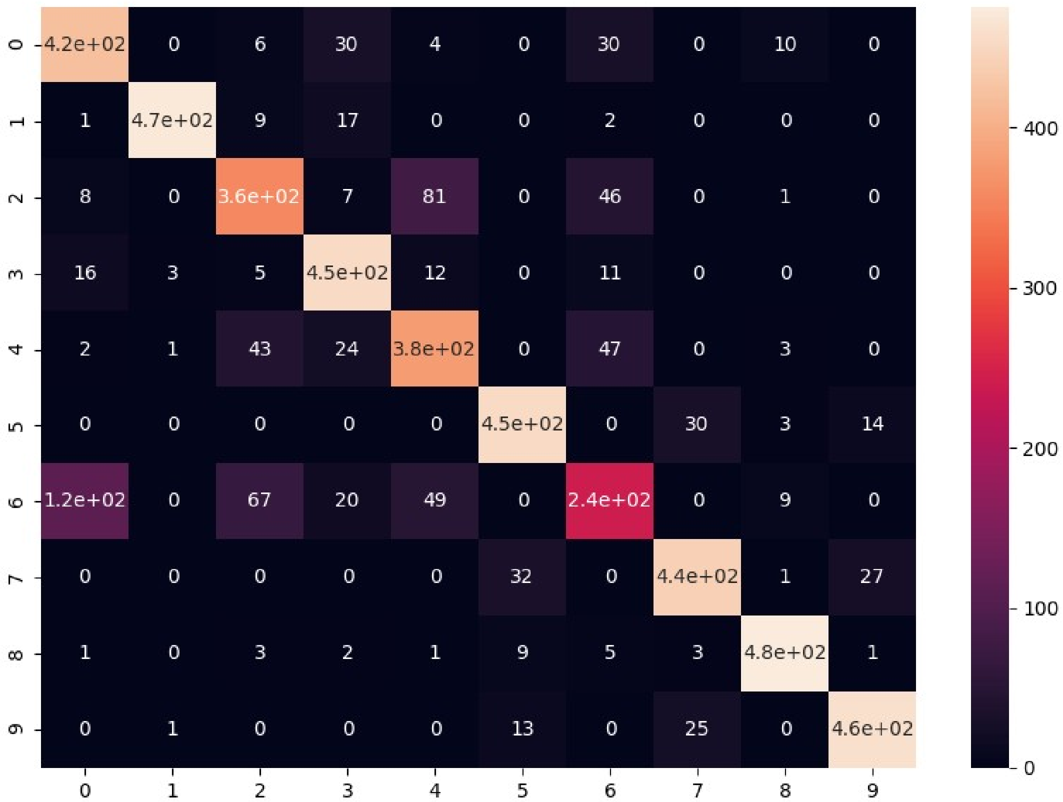
[0 , 3 , 0 , 1 , 0 ,226 , 0 , 10 , 0 , 10],

[64 ,1 ,34 ,0 ,25 ,0 ,123 ,0 ,3 ,0],

[0 , 0 , 0 , 0 , 0 , 38 , 0 ,205 , 0 , 7],

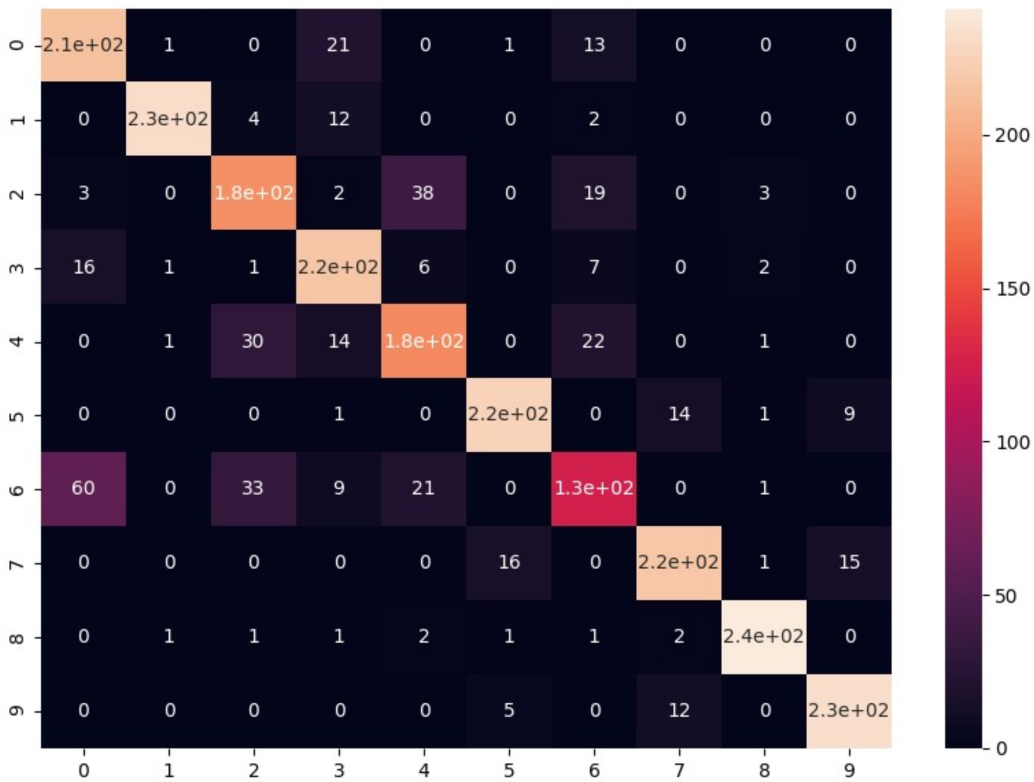
[0 ,1 , 3 ,7 ,5 ,1 ,0 ,2 ,231 , 0],

[0 , 2 , 0 , 1 , 1 ,10 , 0 ,30 , 0 ,206]]



confusion matrix for test data . (using the Scikit SVM library)

```
[[420   0   6  30   4   0  30   0  10   0]
 [  1 471   9  17   0   0   2   0   0   0]
 [  8   0 357   7  81   0  46   0   1   0]
 [ 16   3   5 453  12   0  11   0   0   0]
 [  2   1  43  24 380   0  47   0   3   0]
 [  0   0   0   0   0 453   0  30   3  14]
 [115   0  67  20  49   0 240   0   9   0]
 [  0   0   0   0   0  32   0 440   1  27]
 [  1   0   3   2   1   9   5   3 475   1]
 [  0   1   0   0   0  13   0  25   0 461]]
```



confusion matrix for Validation data .(using the Scikit SVM library)

```
[[214   1   0  21   0   1  13   0   0   0]
 [  0 232   4  12   0   0   2   0   0   0]
 [  3   0 185   2  38   0  19   0   3   0]
 [ 16   1   1 217   6   0   7   0   2   0]
 [  0   1  30  14 182   0  22   0   1   0]
 [  0   0   0   1   0 225   0  14   1   9]
 [ 60   0  33   9  21   0 126   0   1   0]
 [  0   0   0   0   0  16   0 218   1  15]
 [  0   1   1   1   2   1   1   2 241   0]
 [  0   0   0   0   0   5   0  12   0 233]]
```

Article 6 is mis-classified with article 0 most often.

Article 0 is T-shirts , 6 is Shirts .

So the results look to be reasonable.

d.)

The parameter C, common to all SVM kernels, trades off misclassification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly. gamma defines how much influence a single training example has. The larger gamma is, the closer other examples must be to be affected.

REPORT BY :

Shivam Jadhav

2017CS10378