Homework-2

Name: Priyank Devpura **UIN: 225002461**

Compile and Run:

a) **Naïve Bayes:**
   (i) To compile and run the normal Naïve bayes: NaiveBayes.py ../data/imdb1/
   (ii) To compile and run the Filter words Naïve bayes: NaiveBayes.py -f ../data/imdb1/
   (iii) To Compile and run the Boolean Naive Bayes: NaiveBayes.py -b ../data/imdb1/

b) **Perceptron:**

The perceptron was running quite slow when the iterations is very high such as 100 to 1000. So I have randomly sampled the examples for higher iterations.

   (i) To Compile and run: Perceptron.py ../data/imdb1/ <ITER_CT>
   <ITER_CT>= 1,10,50,100,1000
   (ii) To reduce runtime for higher iterations: 100,1000, I have done some sampling of examples:
   Perceptron_100.py ../data/imdb1/ 100
   Perceptron_1000.py ../data/imdb1/ 1000

Results:

	Average Accuracy
Naïve Bayes	0.8165
Filtered Naïve Bayes	0.811
Boolean Naïve Bayes	0.7385
Perceptron Iter=1	0.714
Perceptron Iter=10	0.75
Perceptron Iter=50	0.807
Perceptron Iter=100	0.8225
Perceptron Iter=1000	0.8225

Analysis:

- Among different versions of Naïve Bayes, Normal Naïve Bayes and Filtered Naïve Bayes accuracy is pretty close to each other which implies that filtering removed words that does not affect positive and negative sentiment.
- In Boolean Bayes, we are only focusing on whether the word is present or not. So it does not take into account the multiple count of a particular word that may affect the sentiment. As a result, we are seeing degradation in the output.

- I ran perceptron with different iterations and as the number of iterations increased the accuracy got better. We can see increase in accuracy with increase in the number of iterations as the calculated weights got more accurate.
- Overall Perceptron performed a little better than Naïve Bayes in terms of accuracy.

Limitations:

Perceptron takes a long time if we increase the number of iterations. So in case of higher iterations, I have randomly sampled the examples. Some of the runs take 2-3 hrs with higher iterations in perceptron.

Detailed Results

Normal Naïve Bayes

[INFO] Fold 0 Accuracy: 0.765000
[INFO] Fold 1 Accuracy: 0.850000
[INFO] Fold 2 Accuracy: 0.835000
[INFO] Fold 3 Accuracy: 0.825000
[INFO] Fold 4 Accuracy: 0.815000
[INFO] Fold 5 Accuracy: 0.820000
[INFO] Fold 6 Accuracy: 0.835000
[INFO] Fold 7 Accuracy: 0.825000
[INFO] Fold 8 Accuracy: 0.755000
[INFO] Fold 9 Accuracy: 0.840000
[INFO] Accuracy: 0.816500

Naïve Bayes Filter Words

[INFO] Fold 0 Accuracy: 0.760000
[INFO] Fold 1 Accuracy: 0.825000
[INFO] Fold 2 Accuracy: 0.825000
[INFO] Fold 3 Accuracy: 0.830000
[INFO] Fold 4 Accuracy: 0.800000
[INFO] Fold 5 Accuracy: 0.830000
[INFO] Fold 6 Accuracy: 0.830000
[INFO] Fold 7 Accuracy: 0.835000
[INFO] Fold 8 Accuracy: 0.755000
[INFO] Fold 9 Accuracy: 0.820000
[INFO] Accuracy: 0.811000

Naïve Bayes Boolean

[INFO] Fold 0 Accuracy: 0.700000
[INFO] Fold 1 Accuracy: 0.720000
[INFO] Fold 2 Accuracy: 0.745000
[INFO] Fold 3 Accuracy: 0.735000
[INFO] Fold 4 Accuracy: 0.725000
[INFO] Fold 5 Accuracy: 0.740000
[INFO] Fold 6 Accuracy: 0.755000
[INFO] Fold 7 Accuracy: 0.765000
[INFO] Fold 8 Accuracy: 0.735000
[INFO] Fold 9 Accuracy: 0.765000

[INFO] Accuracy: 0.738500

Perceptron Iteration 1

[INFO] Fold 0 Accuracy: 0.710000
[INFO] Fold 1 Accuracy: 0.730000
[INFO] Fold 2 Accuracy: 0.725000
[INFO] Fold 3 Accuracy: 0.790000
[INFO] Fold 4 Accuracy: 0.675000
[INFO] Fold 5 Accuracy: 0.715000
[INFO] Fold 6 Accuracy: 0.700000
[INFO] Fold 7 Accuracy: 0.645000
[INFO] Fold 8 Accuracy: 0.745000
[INFO] Fold 9 Accuracy: 0.705000
[INFO] Accuracy: 0.714000

Perceptron Iteration 10

[INFO] Fold 0 Accuracy: 0.710000
[INFO] Fold 1 Accuracy: 0.745000
[INFO] Fold 2 Accuracy: 0.740000
[INFO] Fold 3 Accuracy: 0.760000
[INFO] Fold 4 Accuracy: 0.750000
[INFO] Fold 5 Accuracy: 0.745000
[INFO] Fold 6 Accuracy: 0.765000
[INFO] Fold 7 Accuracy: 0.745000
[INFO] Fold 8 Accuracy: 0.765000
[INFO] Fold 9 Accuracy: 0.725000
[INFO] Accuracy: 0.750000

Perceptron Iteration 50

[INFO] Fold 0 Accuracy: 0.805000
[INFO] Fold 1 Accuracy: 0.815000
[INFO] Fold 2 Accuracy: 0.805000
[INFO] Fold 3 Accuracy: 0.810000
[INFO] Fold 4 Accuracy: 0.810000
[INFO] Fold 5 Accuracy: 0.805000
[INFO] Fold 6 Accuracy: 0.830000
[INFO] Fold 7 Accuracy: 0.760000
[INFO] Fold 8 Accuracy: 0.805000
[INFO] Fold 9 Accuracy: 0.825000
[INFO] Accuracy: 0.807000

Iteration 100

[INFO] Fold 0 Accuracy: 0.805000
[INFO] Fold 1 Accuracy: 0.840000
[INFO] Fold 2 Accuracy: 0.830000
[INFO] Fold 3 Accuracy: 0.830000
[INFO] Fold 4 Accuracy: 0.835000
[INFO] Fold 5 Accuracy: 0.825000
[INFO] Fold 6 Accuracy: 0.825000
[INFO] Fold 7 Accuracy: 0.795000
[INFO] Fold 8 Accuracy: 0.810000
[INFO] Fold 9 Accuracy: 0.830000
[INFO] Accuracy: 0.822500

Iterations 1000

[INFO] Fold 0 Accuracy: 0.815000
[INFO] Fold 1 Accuracy: 0.840000
[INFO] Fold 2 Accuracy: 0.790000
[INFO] Fold 3 Accuracy: 0.825000
[INFO] Fold 4 Accuracy: 0.810000
[INFO] Fold 5 Accuracy: 0.835000
[INFO] Fold 6 Accuracy: 0.830000
[INFO] Fold 7 Accuracy: 0.805000
[INFO] Fold 8 Accuracy: 0.825000
[INFO] Fold 9 Accuracy: 0.850000
[INFO] Accuracy: 0.822500