MP 1 Report
Ajit Vijayakumar
4257440

To start off MP 1, I parsed the training files and cleaned the files by removing all punctuation and stop words. The text was all lower cased and then split by white space and this text was all put into a list of lists. The inner lists contained the reviews that we got from the text file, while the outer list was useful because we were able to find out the number of reviews when needed. I generated a vocab list for positive and negative reviews, that was used to count the frequencies in each review. After returning my BOWTrain function, we have the necessary information to compute Multinomial BOW. BOWTrain returns the number of positive and negative words, the number of positive and negative reviews, dictionaries for positive and negative word totals. After calculating the probability for each we can classify the reviews on based on negative or positive.

I got TFIDF which was using 4 dictionaries 2 for IDF and 2 for TF. Compute the TF for each class and then compute the IDF for each class. The function takes in positive and negative reviews.

I spent extra time, which is why I'm resubmitting late with all of the functions completed, to fix my error with Gaussian BOW and TFIDF. In order for me to calculate BOW Gaussian, I had to iterate through positive reviews and sum up the probabilities associated with each review and store them in a dictionary. Then I did the same for negative reviews, but as I iterated through negative reviews, I also computed the number of reviews that we classified correctly in correspondence with positive reviews, which gave me the accuracy of my classifier. I calculated TFIDF very similarly as I used my TFIDF dictionaries and another dictionary to compute probabilities in a similar method as above. At the end of iterating through the positive dictionary, it will start to iterate though the negative dictionary and also start to calculate the reviews that were classified correctly. The function returns accuracy.

The overall run time of my program is about 14 minutes

Multinomial BOW: 73.3% accurate.
Gaussian BOW: 50.0% accurate.
GaussianTDIDF: 50.0% accurate.

Run program with this command:

```
python3 NaiveBayesClassifier.py training_pos.txt training_neg.txt test_pos_public.txt test_neg_public.txt
```