Mohamed Abdulkadir, Ben Walls

In PA3, we built a Sentiment Analysis Classifier for Tweets using naive bayes. Our tasks included the following:

1) Turn training sentences into BOW representations
2) Calculate the conditional probability of each class
3) Calculate the marginal probability of each token under each class
4) Perform estimation using sum of log probabilities (naive bayes approach)

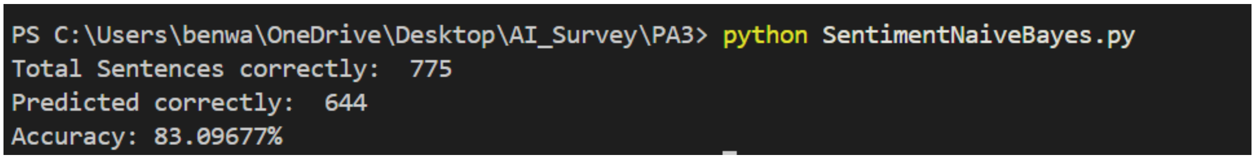
One thing to note about our project is how the marginal probabilities are calculated. On the first attempt, we used a triple for-loop to go through each class, sentence, and word. With the data being quite large, we were concerned about the time complexity of this approach. So, we decided to use matrix multiplication to compute the probabilities. The approach is as follows:

1) Use 3 BOW matrices (one for each class), instead of 1 BOW matrix for all the sentences
2) Remove all columns with all 0s, this makes each matrix contain only BOW representations for sentences of that class.
3) Compute BOW.T x BOW for each class. For each class, the result is a square matrix with side lengths equal to that of the vocabulary length.
4) Take the diagonal from the result of this matrix multiplication. We noticed that since we are using 1s to denote when a token is present in a sentence, the i,i entry of our matrix multiplication for token i is the number of times that token appears in all sentences of that class.
5) Since the diagonal is just the counts of each word, we divide it by the number of sentences in that class, to make a list of marginal probabilities (do this 3 times, one for each class)

Next, we will discuss the use of “epsilon” in our prediction function. We are using this to avoid division by 0 while computing the probabilities. Before adding this, we got some runtime errors, and noticed that math.log(0, 10) is undefined, which is effectively a division by 0. So, for every probability we encountered that had value 0, we made sure to add the epsilon value to avoid this issue.

After we solved this issue, we were able to train and evaluate or model.

Here are the results:



```
PS C:\Users\benwa\OneDrive\Desktop\AI_Survey\PA3> python SentimentNaiveBayes.py
Total Sentences correctly:  775
Predicted correctly:  644
Accuracy: 83.09677%
```