Assignment 3 for Applied Machine Learning 15 Fall

Jingyuan Liu
AndrewId: jingyual
`jingyual@andrew.cmu.edu`

1 Question 1

The modified code is in the amlnaivebayes.py file, and the modified result is in the result.txt file.

From the result.txt file, we can find that, with smoothing, the result is more reasonable. After smoothing, we can find some results are more smoothed, and bigger than 0. Some label scores are 0 before adding smoothing, which is unreasonable because the conditional probabilities should not be 0 for e test case.

2 Question 2

I used SMO and Naive Bayes in weka to do classifications with the dataset. The average precision for SMO is 73.4861%, while the average precision for Naive Bayes is 68.1669%. We can see that SMO is obviously better than Naive Bayes in classification.

I think this is reasonable. Naive Bayes supposes that attributes are conditionally independent given observed label. This assumption may be unreasonbale considering that the features are words. In a doc, words always tend to relate to each other.