

Assignment 1 Report

Zhe Jin email: zj69589@uga.edu

1. The time for the very small data set is about 1.84s, and the accuracy is 3/8.

The time for the small data set is about 258s, the accuracy is 381/818. I run the program on the nike server, nike.cs.uga.edu

2.

3. The Laplace smoothing is helping when the word didn't appear in the training sets, the P may be 0. Using the Laplace smoothing will not effect on the efficiency but will greatly increase the accuracy of the classification.

4. Replacing the Laplace smoothing with a uniform Dirichlet smoother, the efficiency shows little differences, but the accuracy is not as good as the former one.

5. For one way we can consider the multiple features as a new class, so we can do the classification. Or we can use another algorithm like augmented naive bayes, it can show the relationships between the different features.