Bayesian learning for classifying news text articles:

Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents. We will provide a dataset containing 20,000 newsgroup messages drawn from the 20 newsgroups. The dataset contains 1000 documents from each of the 20 newsgroups.

1. Please download the data from http://www.cs.cmu.edu/afs/cs/project/theo-11/www/naive-bayes.html (Newsgroup Data)

2. Please use half data as training data, and the other half as testing data.

3. Please implement the Naive Bayes classifier by yourself (other libraries are not allowed besides basic IO functions). Don't use any online code.

4. Hand in your code and data as a whole package. Please make sure TA can easily run your code.

5. Hand in a report to summarize your implementations and results.

6. Send everything together via canvas.

Due: 11/2/2020