**Assignment 4**

In this assignment you are to create a sentiment analysis classifier using the Natural Language Toolkit classifier module.  As training and test data, use the Movie Reviews Corpus provided by NLTK. Write a short report describing what you did and why. See the sections below for more details on what to include in the report.

**Training**

The training data consists of 2,000 documents (movie reviews) that have been tagged as having negative or positive sentiment.

**Classifier Type**

You will decide which type of classifier to use. Sticking with Naïve Bayes is fine, as is using the code from Ch.6 as a starting point. Experimenting with different classifiers will earn you 5-10 points of extra credit. If you do this, include the code from all the different classifiers you tried, but indicate which system is your final one. Describe what you did in your report and include the results from using each classifier.

**Features**

You will also decide which features to use, but you should experiment with at least two types of features not described in Ch. 6, Sect. 1.3. In your report, describe the features you tried and why you thought they would be useful. Explain which features you kept in your final system and why.

**Evaluation**

Use 10-fold cross validation on the Movie Reviews Corpus, including the results in your report.

**What to hand in and when**

By 11:30 p.m., Thursday, April 28, turn in your code to the Dropbox, along with a pdf of your report.