Mary B. Makarious
COMP 379 — Machine Learning
Saturday 12th Nov. 2016

Homework 4 — Naive Bayes Classifier Write-Up

Data and Pre-Processing:

The goal was to create a binary classifier that would automatically classify movie reviews into positive and negative. The data used was from another repository, but was originally from IMDb, and there were two text files. One file was positive reviews, and the other was negative reviews. Each of the files were read in to their own respective lists, and at the end were merged together, putting the positive reviews at the top and the negative reviews towards the bottom. Each value in the list was then assigned either a positive or negative label depending what was in it. After that, the index values were shuffled around and the data was split into 70% for the training set, 15% for the development set, and 15% for the test set. After analyzing the movie reviews, it seemed as though each review started on a new line. In the code, each review was then read in as a new line and then broken down by words.

Training and Development Sets:

The algorithm implemented was a basic form of the Naive Bayes model. The model was trained by reading in reviews one at a time, taking in their associated labels, and then accounted for each word and the number of times it shows up in the reviews. Testing the model with the development set, after playing around with the number of characters that could be comfortably taken out without hindering performance, it was doing that removing words that were less than two letters long gave good results.

(Going forward, I would try to implement taking out the stop words since that would probably improve performance by taking out words that were already deemed unnecessary…)

Test Set, Results, and Evaluation:

After feeding into the predict function, predictions were made by making a probability score of if the review was positive or negative, and determining which score was the greater of the two. The development set was used for tuning, giving an accuracy of ~76% and an f1 score of ~76% , and evaluating the test set against that gave a similar accuracy of ~75% and an f1 score of ~75%. It was also found that the accuracy scores and the f1 scores were similar, so the evaluation metric used here might not matter as much.

(I personally would trust and go forward with evaluating the f1 score, since it exists to evaluate a test's accuracy in a binary classification. It considers both the precision and the recall of the test to compute the score.)