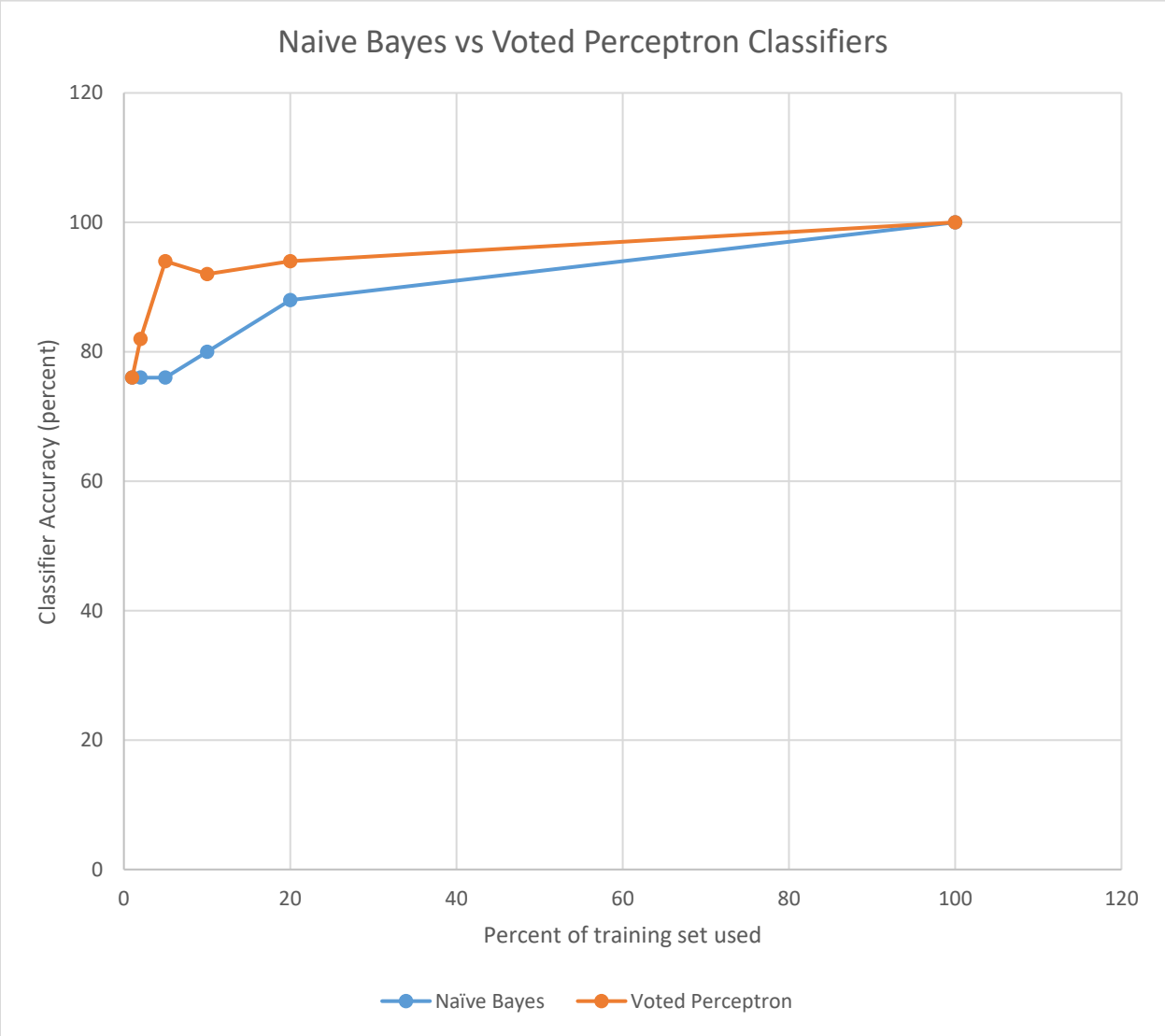
CS 190I: Homework Assignment 2

Abhijit Kulkarni

February 15th, 2017

After creating the Naïve Bayes and Voted Perceptron classifiers, I ran some accuracy tests on them in order to compare their efficiency to see which classifier is better. To do so, I used the last 10% of our training data as the test data; thus it was never tested upon. Then, I tested training the classifiers using 1%, 2%, 5%, 10%, 20%, and 100% of the remaining 90% of data in order to see how well the classifiers predict on different levels of training data. Here is the graph comparing the two.



Here is the table of accuracies for the graph.

Training set %	naïve_bayes	voted perceptron
1%	76%	76%
2%	76%	82%
5%	76%	94%
10%	80%	92%
20%	88%	94%
100%	100%	100%

As you can see, they both predict equally well at 1%. This is because the training size was only 4 documents so it was pretty naïve. However as the size of the training set grew by just a bit, the voted perceptron quickly overtook the naïve bayes classifier in accuracy, while the naïve bayes stayed stagnant until it reached 10%. Interestingly, we see a dip in performance for the voted perceptron at 5%, 10%, 20%. This is probably because the training data there did not match the test set and weights were improperly adjusted. However both classifiers predicted perfectly after they were able to train on a lot of data; this shows that both classifiers are perfectly valid to use, and that even the most naïve can perform very well. Obviously a sophisticated algorithm like voted perceptron works as well, but I was surprised at how well naïve bayes was able to predict the test set.

Note: I used 5 epochs for the voted perceptron.