

CS190I HW2 Report

Matthew Ho

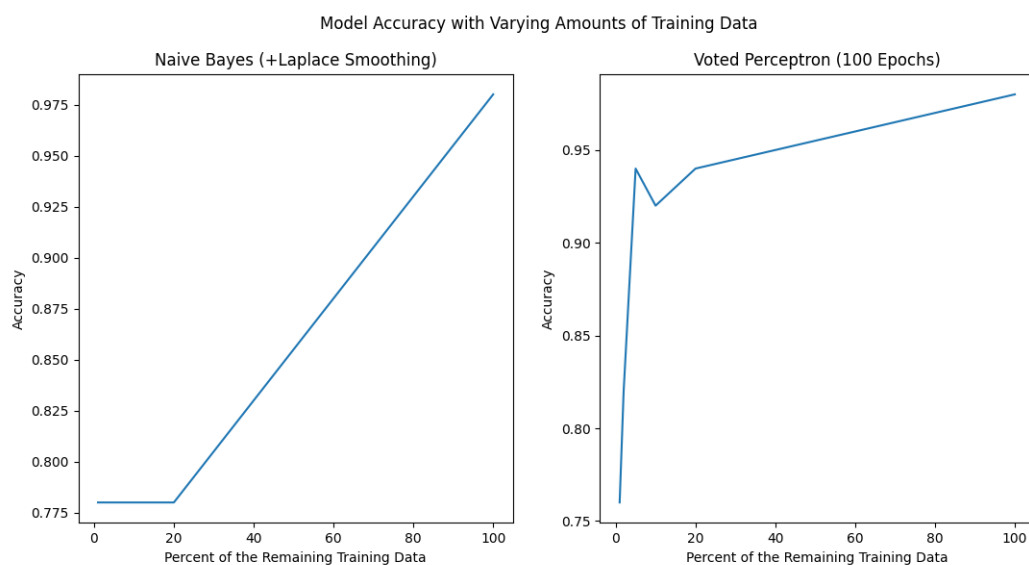
January 2022

1 Collaboration

Did you receive any help whatsoever from anyone in solving this assignment? **No**

Did you give any help whatsoever to anyone in solving this assignment? **No**

2 Report



First of all it should be noted that the last 10 percent of training data is unbalanced (11/50 are fake), so classifiers benefit from being skewed towards guessing non-fake. The Naive Bayes classifier had the same performance for each percentage except for 100. This may be because the maximum likelihood estimates used in the posterior probability computation are more reliable with larger sample sizes.

On the other hand, the Voted Perceptron's performance increased greatly from the 5 initial training examples (1 percent of the remaining data). This makes sense as the model gets significantly more opportunities to adjust its weights from the initial 0 values. The dip from 0.94 to 0.92 accuracy from 5 to 10 percent of remaining training data marks one fewer correct prediction on the $n=50$ "test" set, so I attribute the dip to chance.