CIS520 Report: ML Crackers

Francine Leech, Ziyin Qu, Chen Xiang

December 12, 2016

1 Introduction

2 Preliminary Methods

- eveyrthing tyat didn't work - why it didnt work - image - Ziyin - PCA/GMM - Other ensemble methods methods

3 Main Methods

3.1 Naive Bayes

Accuracy
The second main

3.2 GentleBoost

The second main method we utilized was an ensemble method. We used GentleBoost, a weak learning that was built by MATLAB under the fitensemble function. The method combines many weak learners into one high quality ensemble predictor. We chose this ensemble methods over the others offered by MATLAB, because it is preforms well with binary classification trees with many predictors (ensemble citation).

The input of the model was the $word_train$ data. We used a 10-fold cross validation method to observed how the model preformed, specified the use 300 learners, and the type of learner as 'tree'. The average cross validation error was 0.21. The algorithm classified joy and sadness well.

The method could have improved if we increased the number of learners, however it would have taken a very long time to train because the data is large. Initially we tried the method with the default number of learners, 100 trees, and found that the cross validation accuracy only improved slightly. This slight improvement with triple number of learners reveals that the data has some intricacies or patterns that the ensemble method cannot learn.

3.3 Support Vector Machine

Accuracy
Table with all accuracy

4 Ensemble Method

Ensembl: Sentiment Analysis 1 + Methods

Ensembl: Sentiment Analysis 2 + Methods dictionry wasn't as good

5 Discussion

6 Works Cited

References