Homework Five Report

CMPS242 - Fall 2015

Andrei Ignat Sabina Tomkins Guangyu Wang

1 Intro

2 Data Preprocessing

2.1 Final data used for training

3 Environment

We used python exclusively, and in particular we used scikit-learn for all classifiers.

3.1 Individual Classifiers

Each of the following classifiers we tried on its own, and tried to tune the parameters for. When we combined classifiers we therefore used the best parameters found in the parameter tuning process. The classifiers we used are: Support Vector Machine, Logistic Regression, AdaBoost, Random Forest, Decision Trees, and K Nearest Neighbors. We chose these classifiers as they have been covered in class,

[insert a reason for each one why it may do ok on this dataset]

Accuracy was used as the metric with which to compare classifiers. To see the results from the parameter tuning process please see the Appendix. Both in the parameter tuning, and in the final results below, we used 5-Fold cross validation. However in the ensemble we held out 40% of the training set for validation.

	SVM	Decision Trees	KNN	Random Forest	Logistic Regression	AdaBoost
Accuracy	SVM	Decision Trees	KNN	Random Forest	Logistic Regression	AdaBoost

3.2 Combined Classifiers

4 Results

5 Analysis

6 Appendix

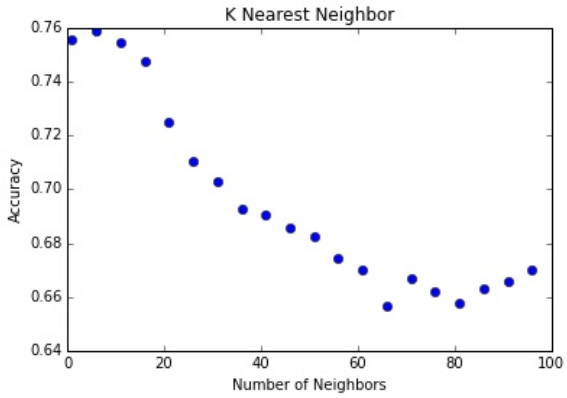

Figure 1: Tuning number of neighbors nearest neighbor

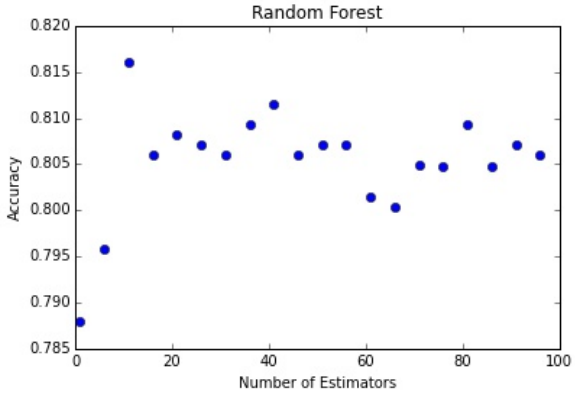

Figure 2: Tuning number of estimators, Random Forest