Jonathan Rudnick

Professor Eric Manley

CS 167: Machine Learning

December 16, 2015

Final Project: Pollster Ratings

**Data**

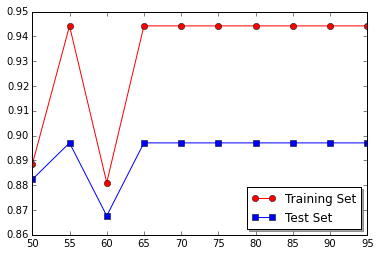
I wanted to combine my interests in Computer Science & Politics for my final project. I generally go to FiveThirtyEight, a site run by Nate Silver, probably the most famous contemporary statistician, for unbiased political news. Silver became famous when he correctly predicted all 50 states in the 2012 presidential election by rating pollsters based on their performance in previous elections. In order to write a pollster rating system of my own, I trained on Silver’s data so that I could predict his pollster ratings.

Before loading the data, I had to edit some characters to make it work with Python. I then used Python to remove some columns that would not be useful for training, changed “Yes” and “No” values to 1’s and 0’s respectively (and the rare “Sometimes” to 0.5), and used dummies to binarize the target column.

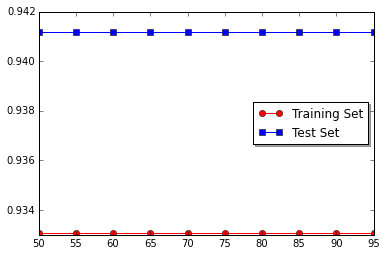
**Methods & Results**

I used SGD Classifier, K-Nearest Neighbor Classifier, SGD Regressor, and K-Nearest Neighbor Regressor to train on the data. For the SGD Classifier, I found that a very low eta0 and a lower number of iterations led to more consistent and more accurate results, generally between 0.868 - 0.956. Using the K-Nearest Neighbor Classifier, I found that, above around 50 neighbors, the results plateaued at 0.897 - 0.956, almost exactly the same results as SGD Classifier. The SGD Regressor worked best with the same parameters as my SGD Classifier, achieving a mean-squared error of 0.072 - 0.077, and the K-Nearest Neighbor Regressor needed somewhat higher neighbors than the Classifier to be consistent at a mean-squared error of 0.043 - 0.055.

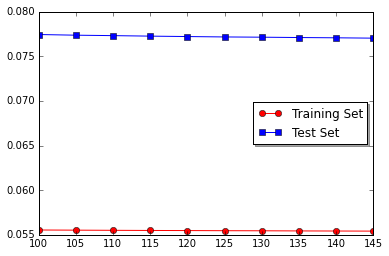
**Figures**



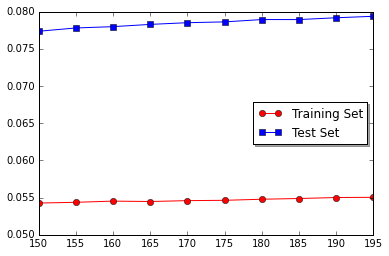
Figure



Figure



Figure



Figure