**Predicting Boston Housing Price**

Kosuke Fukui

**1. Statistical Analysis and Data Exploration**

1.1 Statistics

Number of data points: 506

Number of features: 13

Minimum and maximum housing prices:min 5.0, max 50.0

Mean and median prices of Boston housing prices: mean: 22.53, median: 21.2

Standard deviation: 9.19

**2. Evaluating Model Performance**

I used mean squared error as performance measurement. This error can be divided into Bias and Variance. These values are tradeoff, therefore finding the minimum mean squared error means that finding the best combination of Bias and Variance.

Why is it important to split the data into training and testing data? What happens if you do not do this?

Which cross validation technique do you think is most appropriate and why?

What does grid search do and why might you want to use it?

**3. Analyzing Model Performance**

Look at all learning curve graphs provided. What is the general trend of training and testing error as training size increases?

Look at the learning curves for the decision tree regressor with max depth 1 and 10 (first and last learning curve graphs). When the model is fully trained does it suffer from either high bias/underfitting or high variance/overfitting?

Look at the model complexity graph. How do the training and test error relate to increasing model complexity? Based on this relationship, which model (max depth) best generalizes the dataset and why?

**4. Model Prediction**

Model makes predicted housing price with detailed model parameters

Compare prediction to earlier statistics