

# 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

### 2.1 Performance metrics

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.

### 2.2 Data splitting

I split the data into 336 training data and 167 testing data. If we don't split the data, we can't confirm the generalizing capability of the model.

### 2.3 Cross validation

I used grid search method for cross validation. Grid search is a method to find best parameters for the model. In my program, I used 10-fold cross validation for grid search.

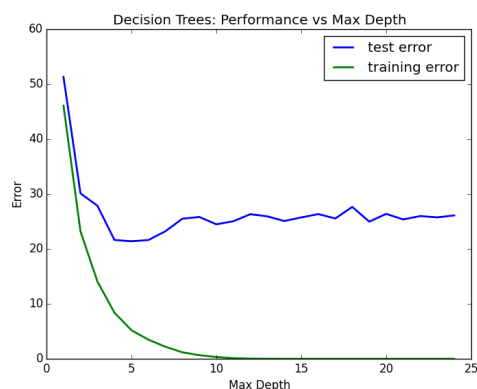
## 3. Analyzing Model Performance

As figure 1 shows, training error increase and test error decrease as training size increase. It means that when the model is fully trained, it is suffered from high bias.



**Figure 1 Performance vs Taraining Size**

As figure 2 shows, training error decrease and test error increase after decrease as model complexity increase. It seems that max depth:5 is the best parameter because sum of the test and training error is the lowest at max depth:5.



**Figure 2 Performance vs Max depth**

#### 4. Model Prediction

Parameters: 11.95, 0.00, 18.100, 0, 0.6590, 5.6090, 90.00, 1.385, 24, 680.0, 20.20, 332.09, 12.13

Prediction Price: 20.76598639

The house price of our client seems to be lower than average price in Boston because the mean of the housing price in Boston is 22.53.