

1) Statistical Analysis and Data Exploration

- Number of data points (houses)? 506
- Number of features? 13
- Minimum and maximum housing prices? $\text{Min} = 5.0$, $\text{Max} = 50.0$
- Mean and median Boston housing prices? $\text{Mean} = 22.5328$, $\text{Median} = 21.2$
- Standard deviation? 9.188

2) Evaluating Model Performance

- Which measure of model performance is best to use for predicting Boston housing data and analyzing the errors? Why do you think this measurement most appropriate? Why might the other measurements not be appropriate here?

Mean squared error. It measures the averages of the squares of the 'errors', that is the differences between the predictions and actual squared error.

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

In order to avoid over fitting, splitting into training and testing able to use the training data to train the algorithm and testing data to test the accuracy.

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

Grid search is a traditional way of performing hyper-parameter optimization. Grid search is an exhaustive searching through a manually specified subset of the hyper-parameter space of a learning algorithm and typically guided by some performance metric such as cross-validation on the training set data or evaluation on a held-out validation set.

- Why is cross validation useful and why might we use it with grid search?

Cross-validation is a model evaluation method that is better than residuals. It performs better than residuals that not require the learner to read all the data.

Grid Search Cross validation implements a 'fit' and 'score' method and used to apply these methods are optimized by cross-validation grid-search over a parameter grid.

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 (max depth) reported using grid search. Note due to the small randomization of the code it is recommended to run the program several times to identify the most common/reasonable price/model complexity.
- Compare prediction to earlier statistics and make a case if you think it is a valid model.