1) Statistical Analysis and Data Exploration

- Number of data points (houses)? 506
- Number of features? 13
- Minimum and maximum housing prices? Min = 5.0, Max = 50.0
- Mean and median Boston housing prices? Mean = 22.5328, 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.
  Mean Absolute Percentage Error, the Mean Absolute Deviation, and the Mean Square Error is appropriate to test for the standard measures. In this case, mean squared error allows us to find the minimum or maximum values, and more accuracy than others on the regression model.

- 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 guised 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 do not require the learner to read all the data. Cross validation solves the limited data by splitting into k smaller sets (Aka k "folds"). By that, all data can be used for training and none has to be held back in a separate test set. That's why it is better than traditional splitting by using this method.

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.

In error analysis, high training error means high bias, which considers as bad and overlapping train/test curves, means no overfitting/variance, which considers as good.

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?

  The learning curve shows that training error is low while testing error is high at the beginning because the model has not been well trained yet.

  The test error decreased as the training size increase. The training error falls steeply on the first few depths and goes close to zero on the 6^{th} depth.

- 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?

  At the first depth, the test error has a sky rocking high error (and falls drastically after reaching to 400 to 50), and the training error has an average 50-error rate consistently. It is under fitting.

  At the tenth depth, the test value lower to 370 compare the 400 from the first depth, and it falls sharply to an average 25 after it reaches to 370. The training error is consistently low that is close to zero on the chart and training error is high. It is overfitting.

  When the model is fully trained, the tenth depth training error has close to zero error

By difference between 23 vs 50 in error, it is obvious the model is over - fitting.

- 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?

  The model complexity grows as the depth grows. However it does not necessary suggested a better model. As the model complexity increases, training error drops significant after the 4th depths and drops to zero on the 10th and test error drops sharply on the first 5th depths. After 5th depths, the testing error decreases while testing error remains the same. However, low training errors lead to overfitting.

  Therefore 5th depths are the best without high computational cost, because it does not over trained the model and it has the least test and training error and also test error at the lowest point.

4) Model Prediction

```
Model Complexity:
Final Model:  GridSearchCV(cv=None,  error_score='raise',
  estimator=DecisionTreeRegressor(criterion='mse',  max_depth=N
one, max_features=None,              max_leaf_nodes=None,  min_sa
mples_leaf=1,  min_samples_split=2,              min_weight_fract
ion_leaf=0.0,  random_state=None,             splitter='best'),
       fit_params={},  iid=True,  loss_func=None,  n_jobs=1,
   param_grid={'max_depth':  (1,  2,  3,  4,  5,  6,  7,  8,  9,  10)},
       pre_dispatch='2*n_jobs',  refit=True,  score_func=None,
      scoring=make_scorer(performance_metric,  greater_is_bette
r=False),         verbose=0)
House: [11.95,  0.0,  18.1,  0,  0.659,  5.609,  90.0,  1.385,  24,  68
0.0,  20.2,  332.09,  12.13]
Prediction:  [ 21.62974359]
Best model parameter:  {'max_depth':  6}
```

The predicted housing price is $21.6 and the average price is $22.5 with a standard deviation of 9.188 and range from 5.0 to 50.0.

Reference:

https://www.quora.com/What-are-good-ways-to-improve-results-while-applying-Machine-Learning-algorithms

http://stats.stackexchange.com/questions/34652/grid-search-on-k-fold-cross-validation

http://jmlr.csail.mit.edu/papers/volume11/cawley10a/cawley10a.pdf