UDACITY

< Return to "Machine Learning Engineer Nanodegree" in the classroom

Predicting Boston Housing Prices

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Your answers show the great amount of effort you have put into learning the concepts. All the best. Happy learning. 👍 😄

Data Exploration

All requested statistics for the Boston Housing dataset are accurately calculated. Student correctly leverages NumPy functionality to obtain these results.

Student correctly justifies how each feature correlates with an increase or decrease in the target variable.

Wow...You have captured the concept of correlation accurately here!
There are many types of correlation in statistics actually. Do you know what kind of correlation you have described in your answer? Check out the following link to find out:
http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

Developing a Model

Student correctly identifies whether the hypothetical model successfully captures the variation of the target variable based on the model's R^2 score.
The performance metric is correctly implemented in code.

Student provides a valid reason for why a dataset is split into training and testing subsets for a model.
Training and testing split is correctly implemented in code.

Analyzing Model Performance

Student correctly identifies the trend of both the training and testing curves from the graph as more training points are added. Discussion is made as to whether additional training points would benefit the model.

Student correctly identifies whether the model at a max depth of 1 and a max depth of 10 suffer from either high bias or high variance, with justification using the complexity curves graph.

Student picks a best-guess optimal model with reasonable justification using the model complexity graph.

Evaluating Model Performance

Student correctly describes the grid search technique and how it can be applied to a learning algorithm.

Student correctly describes the k-fold cross-validation technique and discusses the benefits of its application when used with grid search when optimizing a model.

Great explanation. To reiterate points which you have mentioned:
K-fold CV is an algorithm validation technique: whether a given algorithm will train properly or not. When you get different models from different folds, what you do is average out the evaluation metric of all the models to get what? Well, to get an 'unbiased estimate of model generaliztion on unseen data'. That is the main purpose of k-fold cross validation.

Student correctly implements the `fit_model` function in code.

Student reports the optimal model and compares this model to the one they chose earlier.

Student reports the predicted selling price for the three clients listed in the provided table. Discussion is made for each of the three predictions as to whether these prices are reasonable given the data and the earlier calculated descriptive statistics.

Student thoroughly discusses whether the model should or should not be used in a real-world setting.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review