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.

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.

Excellent! To reiterate the points of this concept:
R-squared value is the percentage of the response variable variation that is explained by a linear model:
R-squared = Explained variation / Total variation
In the target variable, there is some variation contained. The model is trained to learn pattern from the independent variables and try to explain the variation in target variable.
Total sum of squares (SST)= sum of squares explained by regression model(SSR) + sum of squared errors(SSE)
1 = (SSR/SST) + (SSE/SST)
(SSR/SST)=1-(SSE/SST)
R^2 = (SSR/SST) =1-(SSE/SST)
But do you know there are some drawbacks association with R^2 value? If so, then what else should we use in its place as a model evaluation metric? Read through the following link to find out:
http://blog.minitab.com/blog/adventures-in-statistics-2/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables

✓ 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.

Excellent detailed explanation. 😄 Each machine learning algorithm, has parameters and hyper-parameters as their characteristics. Also, the terms parameters and hyper-parameters are not interchangeable. Can you find out what is the difference between them?

✓ 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