UDACITY

PROJECT

Predicting Boston Housing Prices

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Meets Specifications

The project has met all the specifications. Hope you enjoyed working on this project. All the best! 👍 😃

Data Exploration

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

Good job the statistics for the Boston Housing dataset are accurately calculated. To calculate the statistics you could have used numpy functions like for min, max, mean you can use numpy.amin, numpy.amax. numpy.mean etc.
Suggestion- You can refer numpy manual for additional functions

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

You have correctly justified how each feature correlates with an increase or decrease in the target variable
Suggestion- This article reading could be beneficial to further understand correlation.

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.

The performance metric is correctly implemented and the discussion on how successfully the model captures the variation of the target variable is correct

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.

Valid reasoning has been provided as to why a dataset is split into train and test sets. Good job on using random_state as it's very important to reproduce the results
Suggestion - Please refer a very interesting blog on train test split and cross validation

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.

Good job on the discussion, we see that training score decreases, testing score increases (and possibly decreases depending on depth). Adding more points is not beneficial.

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.

As we can see a max_depth of 1 suffers from high bias; a max_depth of 10 suffers from high variance. Well done! the reasoning for the complexity curves is absolutely correct.

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

Good intuition has been provided in selecting the best guess optimal model 

Evaluating Model Performance

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

Basically the grid search technique is such that a set of parameters and values are chosen, and the learning algorithm is run for each combination of parameters to find which model achieves the highest score. Great job on clearly understanding this technique.

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 job on correctly discussing KFold cross validation

Student correctly implements the `fit_model` function in code.

The fit_model has been correctly implemented 
Good to see you have also implemented make_scorer and GridSearchCV too,

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

Excellent work on recognizing the optimal model

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.

The discussion is apt, Client 2 has a low prediction, Client 1 a mid prediction, and Client 3 a high prediction. The predictions do seem reasonable.

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

Perfect reasoning has been given as to why the model is not appropriate in a real world setting!

DOWNLOAD PROJECT

RETURN TO PATH

Student FAQ