UDACITY

Predicting Boston Housing Prices

REVIEW

CODE REVIEW

HISTORY

Requires Changes

4 SPECIFICATIONS REQUIRE CHANGES

Hello Udacian,

I am sure you must have worked on this for hours and you should be proud of your work.

You have successfully passed most of the rubrics of this project excellently on the very first submission. There are only 4 specifications that require some changes, it will hardly take half an hour, as you just need to add few lines which were an important part of the rubric.

Please see my comments below and provide the answers.

I wish you 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.

You need to use the NumPy functionality to obtain these results.

Here is an example:

import numpy as np

a = [1,2,3,4,5,6]
minimum= np.min(a)

Please checkout this link to obtain the result :
https://docs.scipy.org/doc/numpy-1.14.0/reference/generated/numpy.minimum.html

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

Well Done!
You have correctly defined each feature and the relationship with 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.

Great!
You have correctly identified whether the hypothetical model successfully captures the variation of the target variable based on the model's R^2 score.
You have correctly implemented the performance metric here.
Suggestions and Comments:
You can checkout this link:
https://www.datasciencecentral.com/profiles/blogs/regression-analysis-how-do-i-interpret-r-squared-and-assess-the

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.

Training and testing split is correctly implemented in code.
Good implementation of random_state for the train_test_split function.
You have provided a valid reason for why a dataset is split into training and testing subsets for a model, just explain how we can avoid the issue of overfitting using test train split method.
Suggestions and Comments:
You can checkout this link:
http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6

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.

Well done!

You have correctly identified the trend of both the training and testing curves from the graph as more training points are added. And you have made your point about adding more training points would benefit the model or not.

Good job describing how the training and testing score change as the training set size increases.

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.

Good job identifying that the model suffers from high bias when max_depth = 1 and that the model suffers from overfitting (high variance) when the max_depth = 10.

Suggestions and Comments:

Bias occurs when a model has enough data but is not complex enough to capture the underlying relationships. Simply put, bias occurs when we have an inadequate model (underfitting), and variance is a measure of how much the predictions vary for any given test sample. High bias gives high error on training set and high variance give high error on testing set.

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

Good job!!

Evaluating Model Performance

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

You got it correct. Grid search helps to find the best possible hyper parameters of the model by applying the exhaustive brute force search. It works good with most of the machine learning models.

Suggestions and Comments:

You may want to check out this link:

http://scikit-learn.org/stable/modules/grid_search.html#grid-search

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.

Please provide a detailed description of k-fold cross validation and how it is performed on a model.

Suggestions and Comments:

You may want to check out this link:

http://scikit-learn.org/stable/modules/cross_validation.html

Student correctly implements the `fit_model` function in code.

You have correctly implemented the fit_model function in your code using the make_scorer and GridSearchCV.
Good Job!

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.

Good Job!
You have correctly predicted the selling price for the three clients.
And you have very well explained the reason behind these prices for each client.

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

I agree!!
But you can improve your answer here, try to answer all Hint questions one by one.

☑ **RESUBMIT**

⬇ **DOWNLOAD PROJECT**

Learn the best practices for revising and resubmitting your project.

RETURN TO PATH

Student FAQ

RETURN TO PATH