Caution: Don't touch your test-dataset at all unless you are absolutely sure about your model. Once you are absolutely ready, then use the predict() method, otherwise you will compromise your test_dataset.

Lab 2-Part A

In this lab, you will work on the same housing data problem. Please follow the steps below:

Optional: If you want, you can apply feature scaling. Then observe how it changes your MSE and r2 score, and write a discussion on that. Feature Scaling is part of your data preprocessing step.

Optional: If you are interested, check out the RANSACRegressor class to deal with outliers and apply that to your model.

[1] First, you need to perform all the data preprocessing parts. Then, you need to select one feature that has a linear relationship with your target variable. Please discuss or show your work why you have selected this particular feature to build your **simple linear regression model**.

[2] Train your model using the fit() method.

[3] Print out the parameters value.

[4] test your model using the predict() function.

[5] Compute the MSE of your training and test predictions. Write a discussion on these values (such as what's your opinion, it's overfitting or underfitting).

[6] Compute the R^2 value of your training data and test data. Write a discussion on these values (such as what's your opinion, it's overfitting or underfitting).

[*** you need to include `from sklearn.metrics import r2_score` and `from sklearn.metrics import mean_squared _error` to work with them. Also, read about the definition of these performance metrics]

Lab 2 – Part B

Optional: If you want, you can apply feature scaling. Then observe how it changes your MSE and r2 score, and write a discussion on that.

Optional: If you are interested, check out the RANSACRegressor class to deal with outliers and apply that to your model.

[1] Use all/some the features to build your **multiple linear regression model**. Discuss why you chose some features/all features.

[2] Train your model using the fit() method.

[3] Print out the parameters value.

[4] test your model using the predict() function.

[5] Compute the MSE of your training and test predictions. Write a discussion on these values (such as what's your opinion, it's overfitting or underfitting).

[6] Compute the R^2 value of your training data and test data. Write a discussion on these values (such as what's your opinion, it's overfitting or underfitting).

[*** you need to include from sklearn.metrics import r2_score and from sklearn.metrics import mean_squared _error to work with them. Also, read about the definition of these performance metrics]

[7] Compare your performance metrics with simple linear regression model and write a discussion.

Lab 2 – Part C

In this part, you will model the non-linear relationship in the housing data. You will model the relationship between house prices and and LSTAT using second-degree (quadratic) and third-degree (cubic) polynomials and compare that to a linear fit. You will first create a regressor object from the LinearRegression class. Also, you will be using PolynomialFeatures transformer class for this part.

[1] Create quadratic/cubic features such as, also you need to use the fit_transform() method.

https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html

>>quadratic = PolynomialFeatures(degree =2)

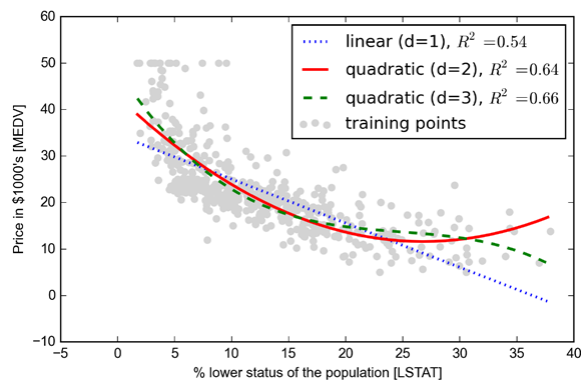……………………

………………………….

[2] Scale features using MinMaxScaler()

https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

[3] train, predict and find out the r2 score for the linearfit

[4] train, predict and find out the r2 score for the quadraticfit

[5] train, predict and find out the r2 score for the cubicfit

Generate a plot like below:

[6] Write a discussion on these three different fits and share your opinion on which fit is the best.

Lab Help:

To learn about how a sklearn class works, please go to the sklearn website and go over the examples and discussion.

What to submit?

You should create 3 .ipynb files for 3 parts and make a zipped file and submit.