

CSST 102

Mangulabnan Jasper
BSCS 3A

As I review the data file I found that it contains information about the house price, including its feature like the bedrooms, size in square feet and the age of the house and according to the instruction given, I/we have to predict the price of the house based on its features.

The dataset is loaded with the use of panda DataFrame (`pd.read_csv`). As I input the function to find the missing value, it revealed that there are no missing values on the format or in the data, although there are no issues nor missing values in the data, further preprocessing is necessary to normalize the features and handle missing values.

To ensure that the feature is in similar scale, the standard scaler is use. This technique allows the feature to transform to a range between 0 and 1, this technique is effective as it helps for faster converge and better performance.

Using python, I make a simple linear regression model without relying on pre-built libraries as said on the instruction. The model was based on LQM or least square method to derive the model parameter or the slope and intercept.

Then the linear regression model was made with method to fit the model (`fit`) and to make predictions (`predict`).

As said on the instruction to split the dataset, I split into two which is the training and the testing with the use to 80-20 split. This work helps the model to evaluate the unseen data.

The model was trained using the training set by fitting it to the data, so this involved calculating the coefficient using normal equation.

The MSE was computed for both training and testing data to evaluate the model's fit. The MSE is used to measure the average squared difference between actual values to predicted values.

The trained model was evaluated on the testing set by making predictions and calculating the MSE.

To visualize the findings, a plot was made showing the regression line along with the test data points. This visual aid helps to understand how well the model fits the data and Identifies any discrepancies between predicted and actual values.

In conclusion, the linear regression model helps for a deeper understanding of its workings and evaluation metrics. The model's performance was assessed using the MSE on both training and testing data. The results provided insights into how well the model generalizes to new data.

The challenges encountered are normalization and model complexity. In normalization, ensuring all features were on a similar scale was crucial. The Standard Scale was used to address this challenge effectively.

Then in Model Complexity, implementing the model from scratch required careful management of matrix operations and understanding of the normal equation. Ensuring the correctness of these operations was essential for accurate results.

In overall, doing this project is an exercise to understand the fundamental machine learning concept and evaluating performance