Attendance: 10%, Continuous evaluation: 70%, Viva-20%

Assignment No. 1

i.   Download House Prices Data Set (already in the needed format). The data set is used to predict house prices. Analyze the columns of the dataset.

Upload the Dataset in the "ML_DRIVE/Assign_1" folder, if executing through COLAB. Access the dataset from there.

ii. Read the dataset in the Pandas data frame. Remove the rows with a missing value. Divide the training.csv into two sets of ratio 80:20 entitled to train and test set respectively.

iii. Use the linear regression method to estimate the slope and intercept for predicting 'SalePrice' based on 'LotArea'.

iv. Use the multiple regression method to estimate the value of the weights/coefficients for predicting 'SalePrice' based on the following features:

Model 1: LotFrontage, LotArea
Model 2: LotFrontage, LotArea, OverallQual, OverallCond
Model 3: LotFrontage, LotArea, OverallQual, OverallCond, 1stFlrSF, GrLivArea

v.   Calculate and compare the Mean squared Error, R2 score for each of the model for test and training set for the above models.

vi. Use the multiple regression method to estimate the value of the weights/coefficients for predicting 'SalePrice' based on the following set of mixed (numerical and categorical) features:

Model 4: LotArea, Street
Model 5: LotArea, OverallCond, Street, Neighborhood
Model 6: LotArea, OverallCond, Street, 1stFlrSF, Neighborhood, Year

vii. Compare the feature "LotArea" weights/coefficients for all the six trained models and plot a graph using the Matplotlib library.

viii.   Use the polynomial regression of degree (2 and 3), to estimate the value of the weights/coefficients for predicting 'SalePrice' based on 'LotArea'. Print the graph on the training and test set (Bonus).

Submit a report with the result.