Information Engineering and Technology Faculty
German University in Cairo

Machine Learning Assignment 1

Name: Youssef Tarek Mohamed
ID: 40-16513

Table of Contents

Assignment 1 Machine Learning

"Using Train, Cross Validation & test dataset"

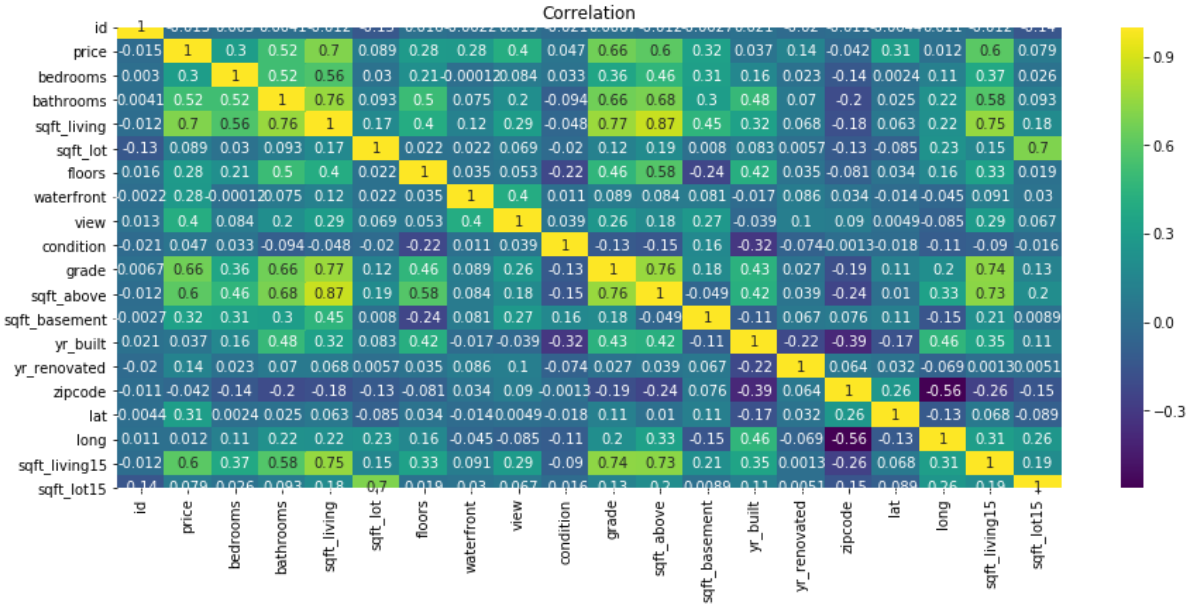
Importing & Dividing Data set

The data is imported using pandas, using pd.read_csv, then data is divided into three data sets "Train" with 60% of the original dataset, "Validate" with 20% of the original dataset and "Test" with 20% of the original dataset.

"Train" is then divided into XTrain & YTrain in the XTrain the selected features will be saved and in the YTrain the Price will be saved.

Same is done with the Validate and the Test they're divided into XValidate, YValidate, XTest and YTest.

Heatmap of Data set features



Heat map is used here to know the most important features presented in the data set and the most effective to be chosen.

Gradient Descent & Compute Cost

Gradient Descent & compute cost are used to return and show us the value of the Theta and the degree we're dealing with, after multiple of trials we can compare each cost and degree to reach the minimum needed cost we do that by changing the selected features that was selected from the heat map. We then use the 20% of the Cross Validation to reach the minimum cost and make sure that same degree is used as train set, after that we should try using the theta we reached and that degree with the XTest and YTest and make sure that the minimum cost resulting is almost the same as the one we reached by Train & Cross validation sets.