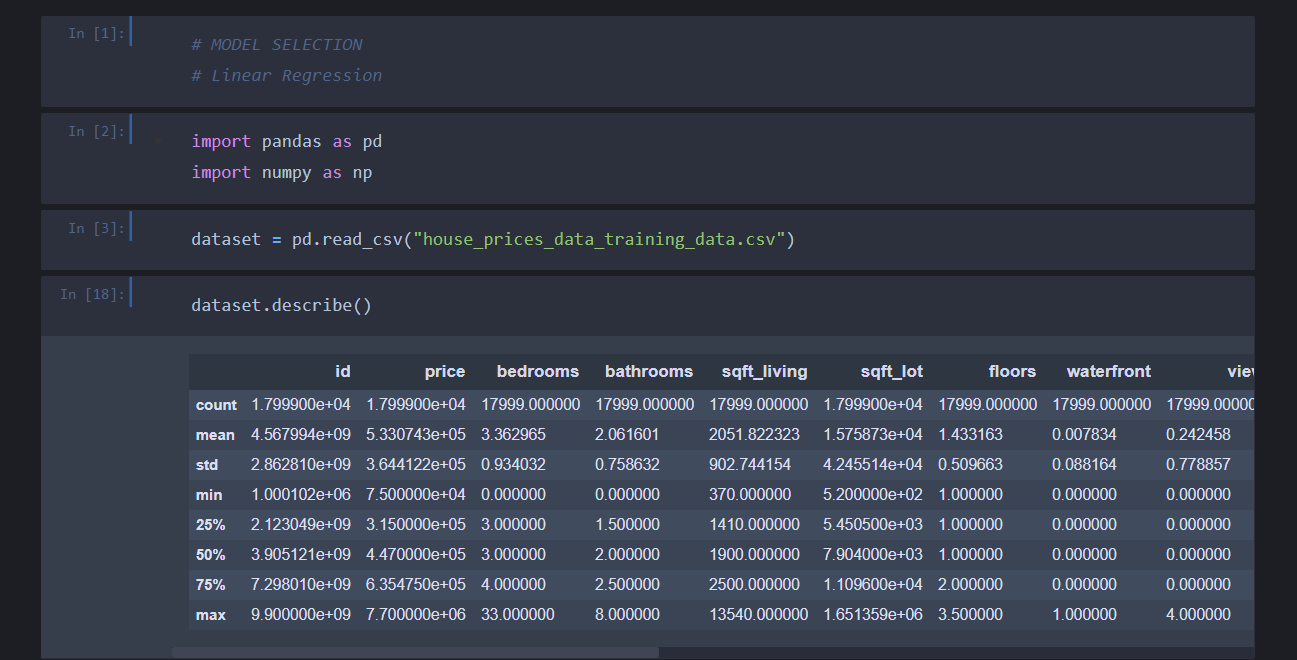
Linear regression

Name : Nardine wael

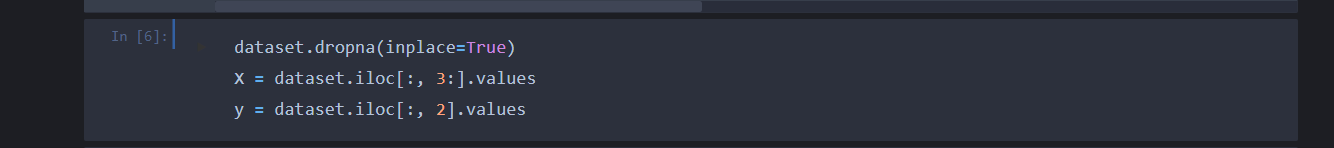
Id :40-8622

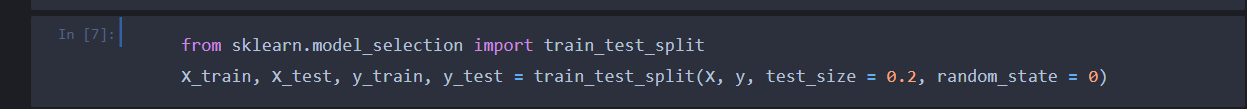
Under the supervision of Dr: Maggie

I had done the code two times one with using scikit learn, and another one without libraries

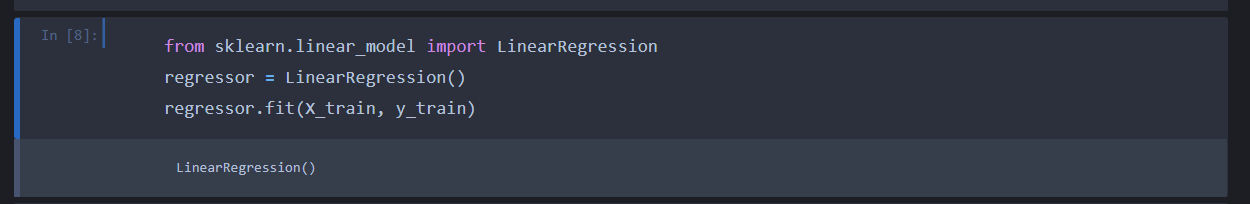
The first time using sci-kit learn library

first, we reading the csv using pandas library. Then I am using the pandas describe method to explore the data set .



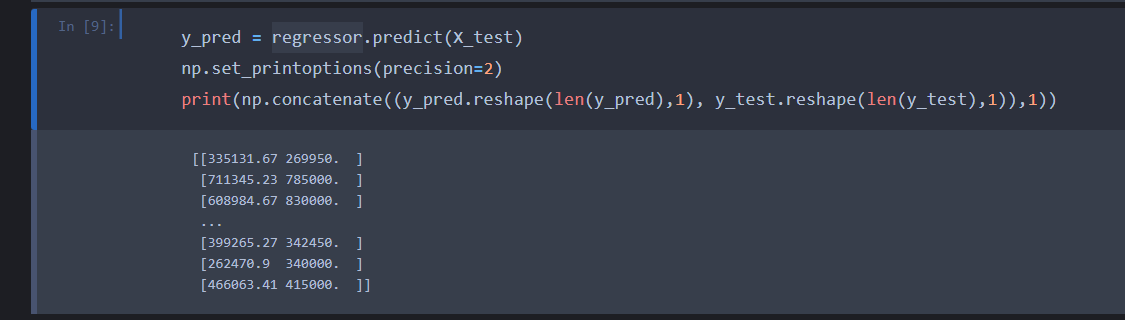
We exclude the rows with the null values .then I am separating my datasets into x and y. x means the dataset features( number of bedrooms ,bathrooms,etc ---).y means the price of the house . 

Here I splitted the data into training and testing dataset .where in the training I will train my model and using the testing data I will test my model’s accuracy .

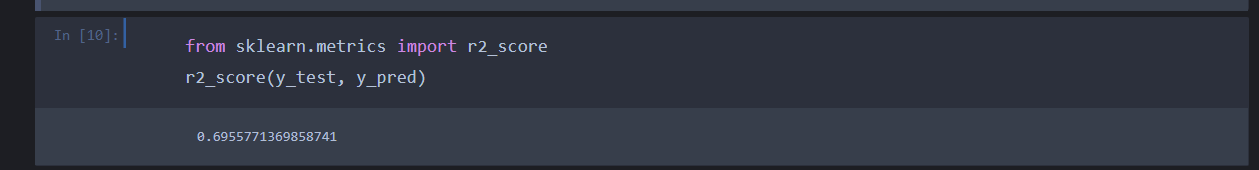


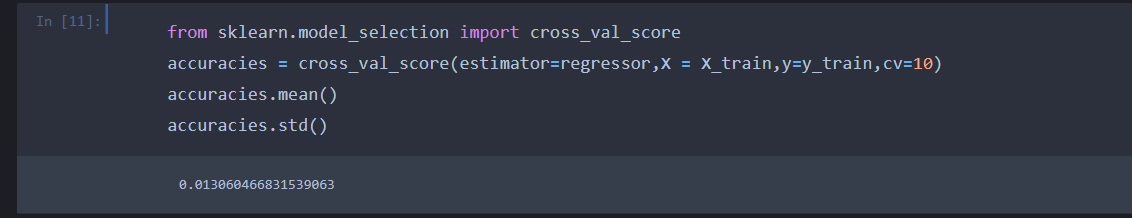
I am initializing my model in this screenshot .My model is linear regression without regularization .

I am training my xtrain data using y train .

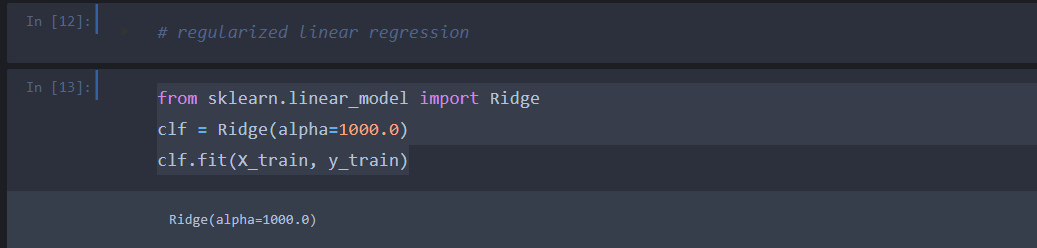


I am trying to predict the y values from the x test values .



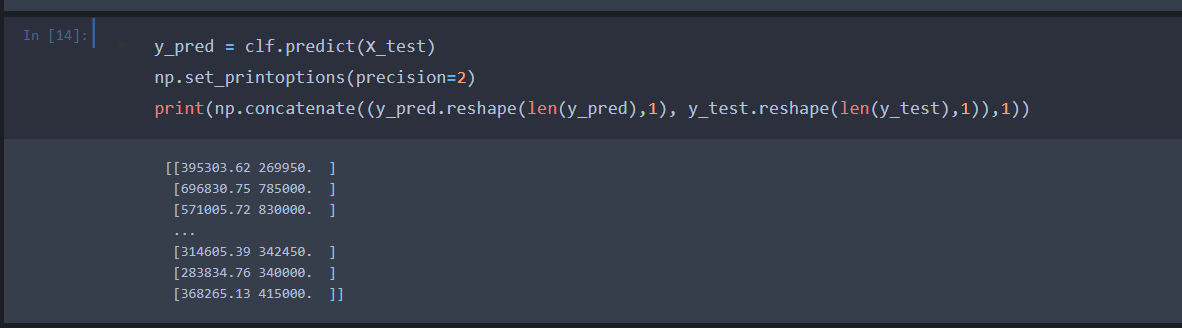
From the predicted y values using the x test I am comparing the predicted values with the original y test values then, based on this I am giving my model accuracy score .

Using the cross validation , I am calculating the model accuracy , accuracy mean, the accuracy standard deviation.

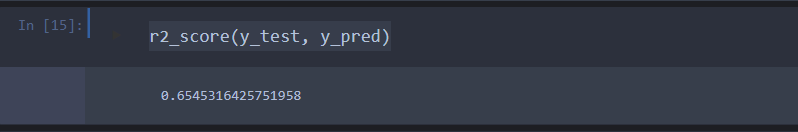


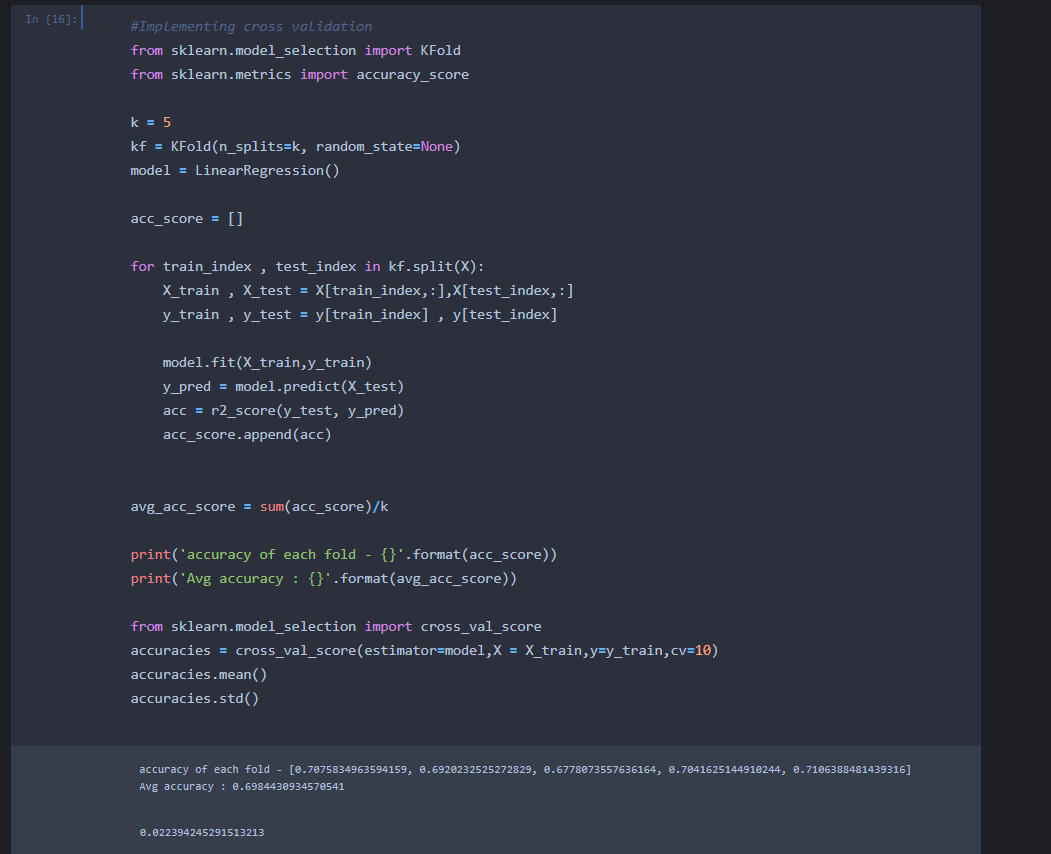
I am initializing my model in this screenshot .My model is regularized linear regression .then

I am training my xtrain data using y train .

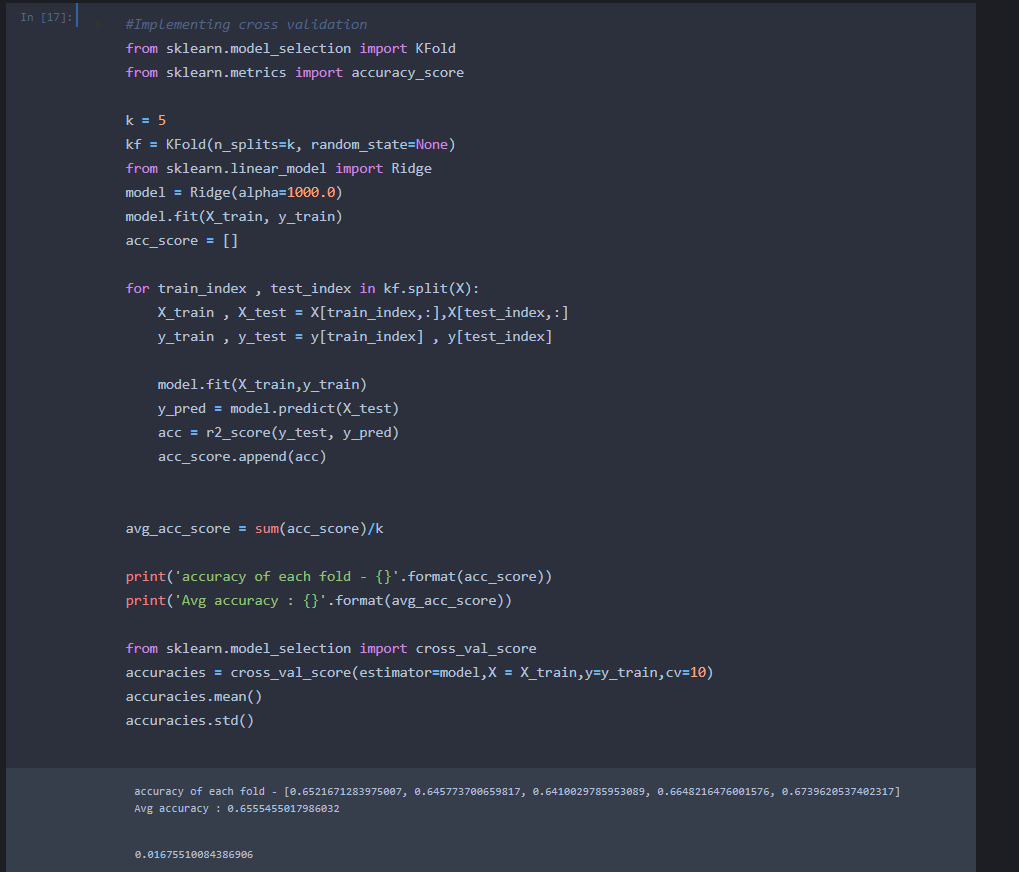


I am trying to predict the y values from the x test values .

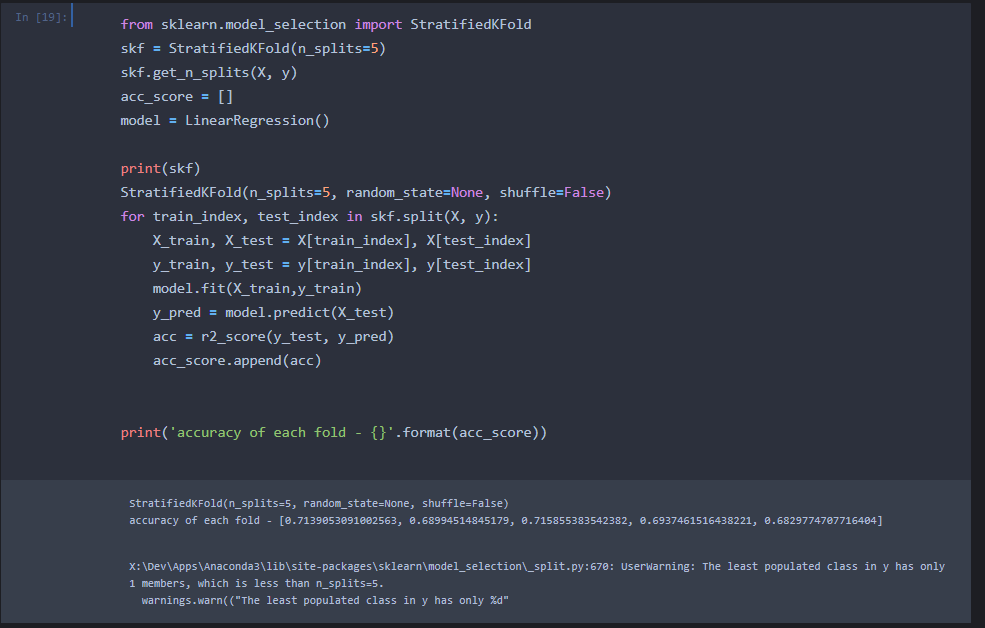


From the predicted y values using the x test I am comparing the predicted values with the original y test values then, based on this I am giving my model accuracy score .

Using k fold I am separating my data into 5 sections then I am splitting my data again using this 5 different section and I am training my data using the new 5 sections and then I am calculating the accuracy for this 5 sections of data .

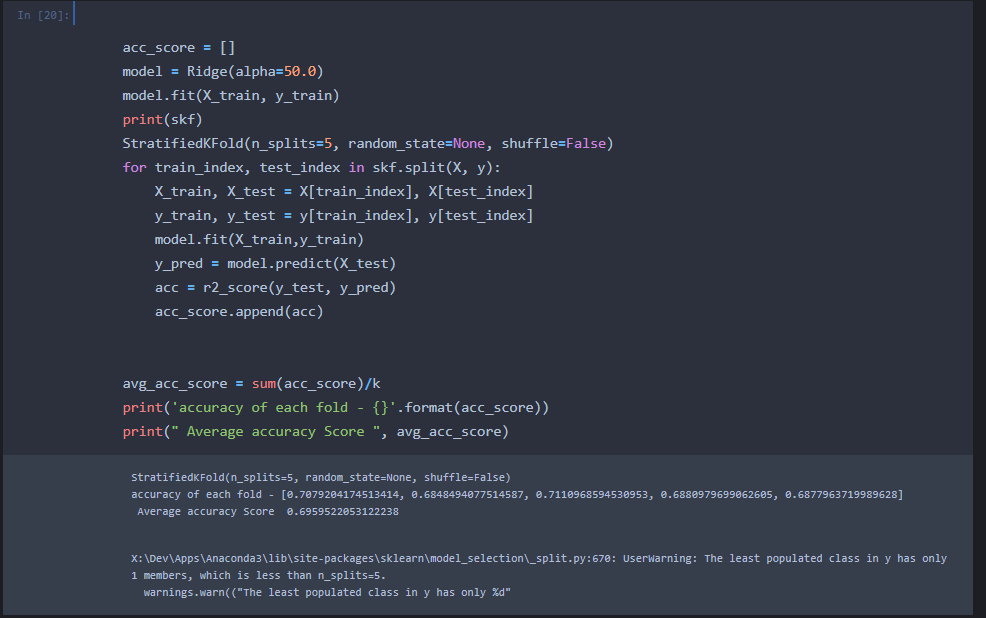


Using k fold I am separating my data into 5 sections then I am splitting my data again using this 5 different section and I am training my data using the new 5 sections and then I am calculating the accuracy for this 5 sections of data using regularized linear regression model .



Allows you to break data into train and test sets using train/test indices.

This cross-validation object returns stratified folds and is a variant of KFold. The folds are created by keeping the percentage of samples for each class constant using linear linear regression.



Allows you to break data into train and test sets using train/test indices.

This cross-validation object returns stratified folds and is a variant of KFold. The folds are created by keeping the percentage of samples for each class constant using regularized linear regression .

The second time of the code without library :

Text

Description automatically generated

First,we import the numpy and pandas library then I am reading the csv into a variable and then I am dropping all the rows with missing values and dropping the first 2 columns the id and date because they are useless. I am converting the data types of my dataframe into float .

Text

Description automatically generated

In feature normalize function, I bring every value in all columns and then I am subtracting the mean of the column and then I am dividing it by the standard deviation of this column to normalize my data.

Text

Description automatically generated

I am reading here the output of the normalization function .

Graphical user interface, text

Description automatically generated with medium confidence

Cost function is the Root Mean Squared Error (RMSE) between predicted y value (pred) and true y value (y\_test). MSE measures the average squared difference between an observation’s actual and predicted values. The output is a single number representing the cost, or score, associated with our current set of weights. Our goal is to minimize MSE to improve the accuracy of our model

Gradient Descent

Text

Description automatically generated

Gradient descent is an optimization algorithm that's used when training a machine learning model. It's based on a convex function and tweaks its parameters iteratively to minimize a given function to its local minimum

A gradient simply measures the change in all weights with regard to the change in error. You can also think of a gradient as the slope of a function. The higher the gradient, the steeper the slope and the faster a model can learn. But if the slope is zero, the model stops learning. In mathematical terms, a gradient is a partial derivative with respect to its inputs.

Text

Description automatically generated

I am adding to the x data frames dummy variable that will be multiplied by the bias of theta .then I am initializing my theta with zeros with the same length of x columns .

Text

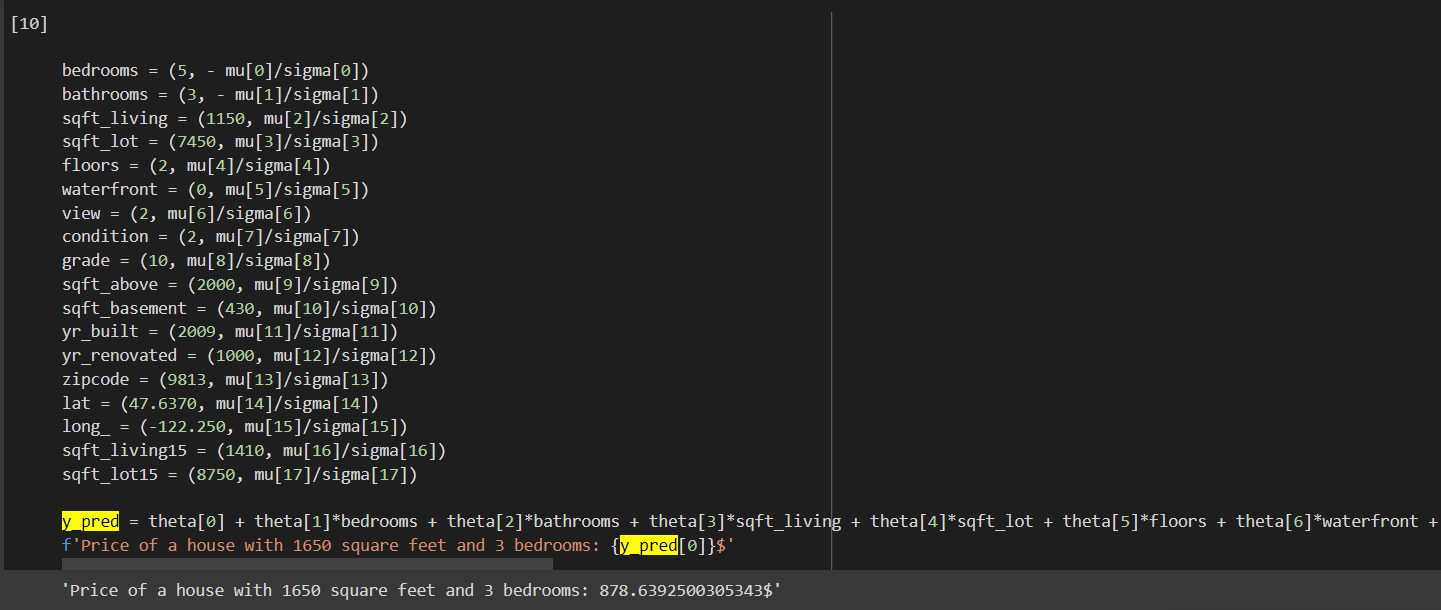
Description automatically generated

Alpha is the learning rate that will be used in my gradient descent. I am trying different values to find the best one to get the lowest local minimum in the gradient descent and the number of iterations = 50.

Graphical user interface, text

Description automatically generated

Using the gradient descent function and it’s outputs and different learning rates ,we will try to find the best alpha .from the graph the best alpha is 0.3.

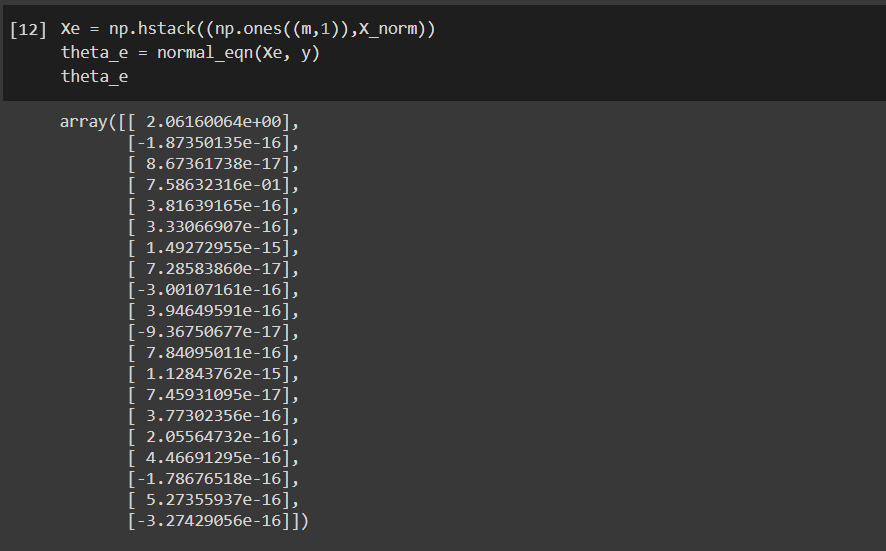


I am trying my output and predicting the price of a custom house .

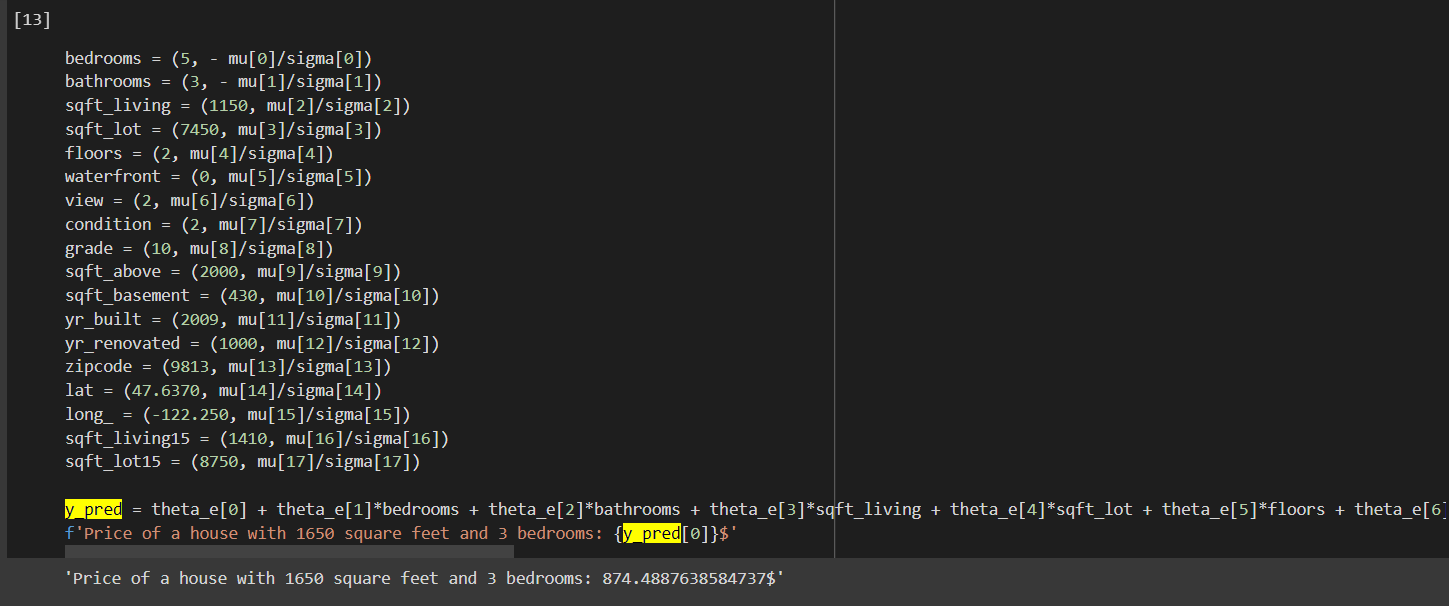
Text

Description automatically generated

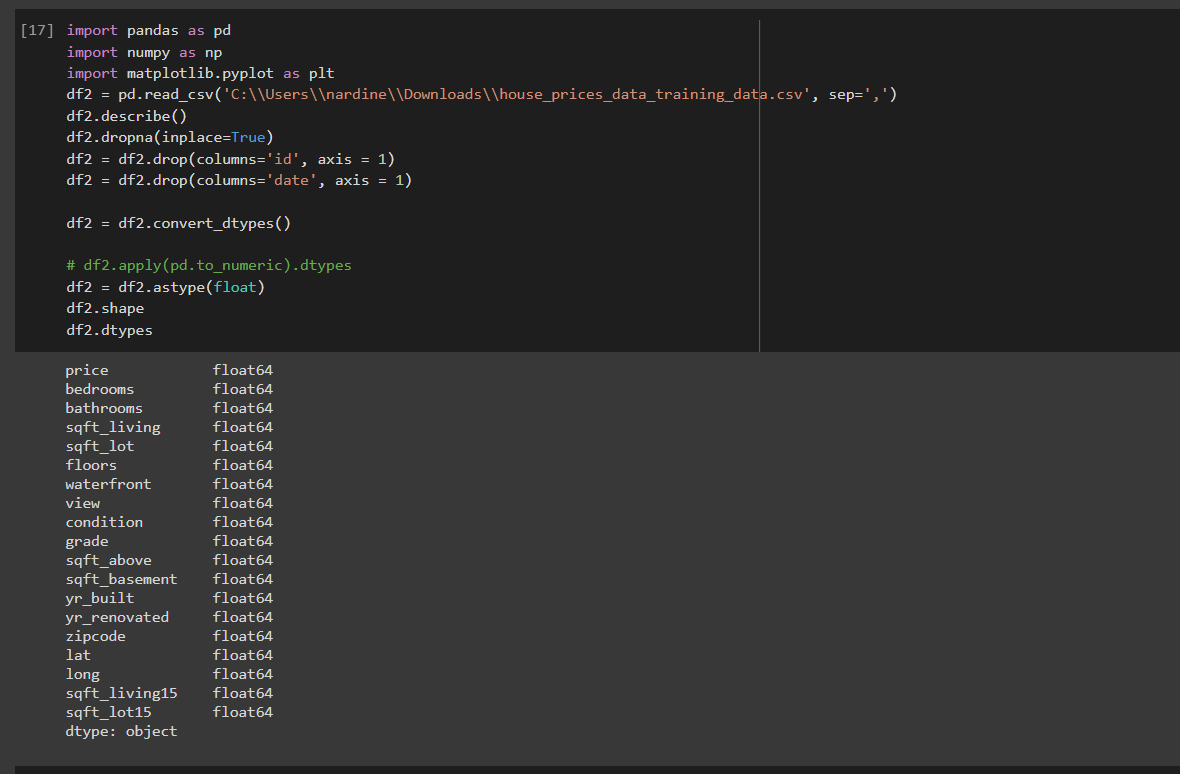
We are solving the linear regression using the normal equation method .



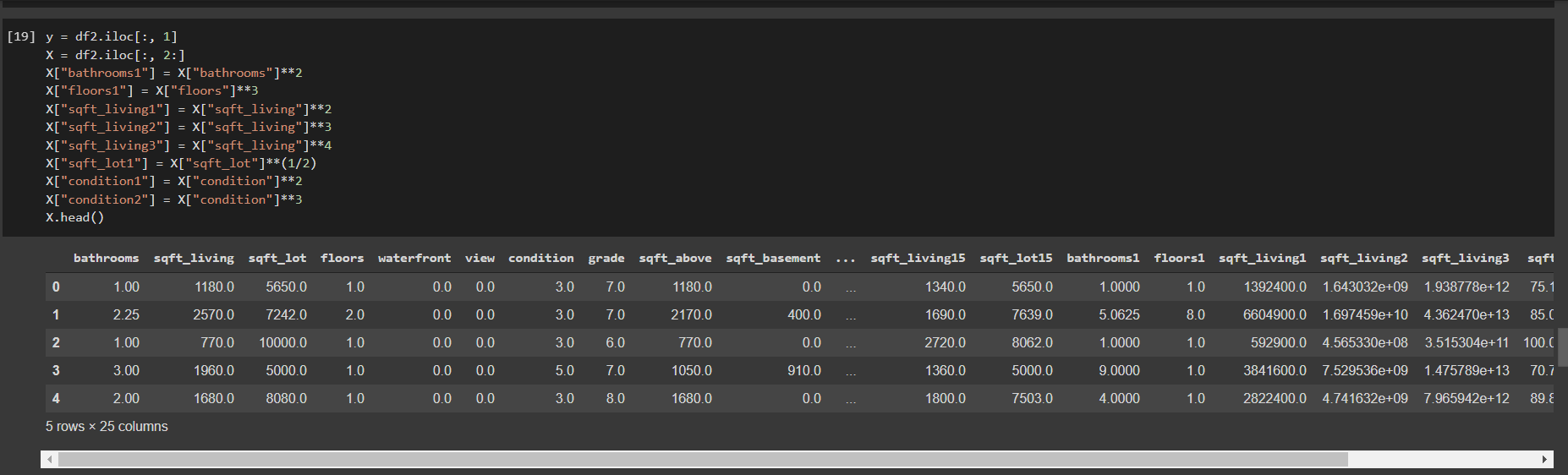
I calling the normal equations and printing the thetas .



I am using the output theta to predict the price of a custom house .



I am re -initializing my dataset to use it in the polynomial linear regression then I am dropping all the nan rows and dropping the ID and Date columns and converting the datatframe dtype to float



I am using a polynomial of degree 4 in this case .

Text

Description automatically generated

Using the model selection, I am splitting my data into 3 things test dataset ,training dataset,cross validation dataset.

A picture containing text, screenshot, indoor, several

Description automatically generated

Using the training dataset I am using different polynomial model with a degree of 4.