Assignment1 Report

Which type of Linear regression is applicable for which data set:-

DataSet	Linear Regression Applicable	Type of Linear Regression
Data1	Yes	Standard Linear Regression
Data2	Yes	Standard Linear Regression after non linear transformation
Data3	No	
Data4	Yes	Multiple Linear Regression

Different Error function's value of my own model:-

DataSet No.	Mean Square Error	Mean Absolute Error	Root Mean Square Error	R Square
Data1	34.6208083	5.15550563	5.88391713	0.98417491
Data2	8892037637.21270	25883.63652443065	94297.60143934045	0.7882858000815468
Data3	0.16173044	0.29467793	0.40215723	0.31369732
Data4	2.0785254	1.28055598	1.44170919	0.95795719

Different Error function's value of Sklearn model:-

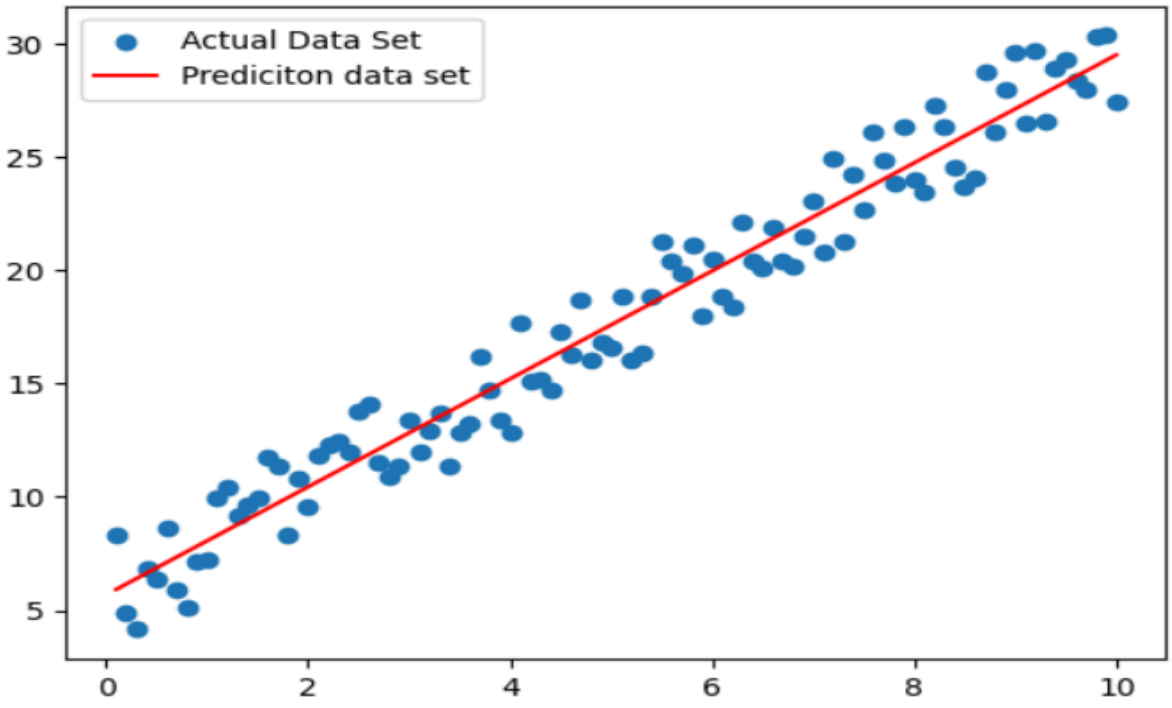
DataSet No.	Mean Square Error	Mean Absolute Error	Root Mean Square Error	R Square
Data4	34.62048082924355	5.155505630378592	5.883917133104744	0.9841749058943147
Data2	8892037637.212702	25883.636524430647	94297.60143934045	0.7882858000815467
Data3	0.16173044143088558	0.2946779330131047	0.4021572347116058	0.3136973226728078

Data4	2.078525401777327	1.280555978429147 6	1.441709194594155	0.9579571905586357

I have used Multiple Linear Regression for all data sets. For all dataset I have added one more column as “X0” which is nothing “C” or intercept in y =mx + c. While adding this column I have taken its weight or coefficient as 1.

DataSet1:-

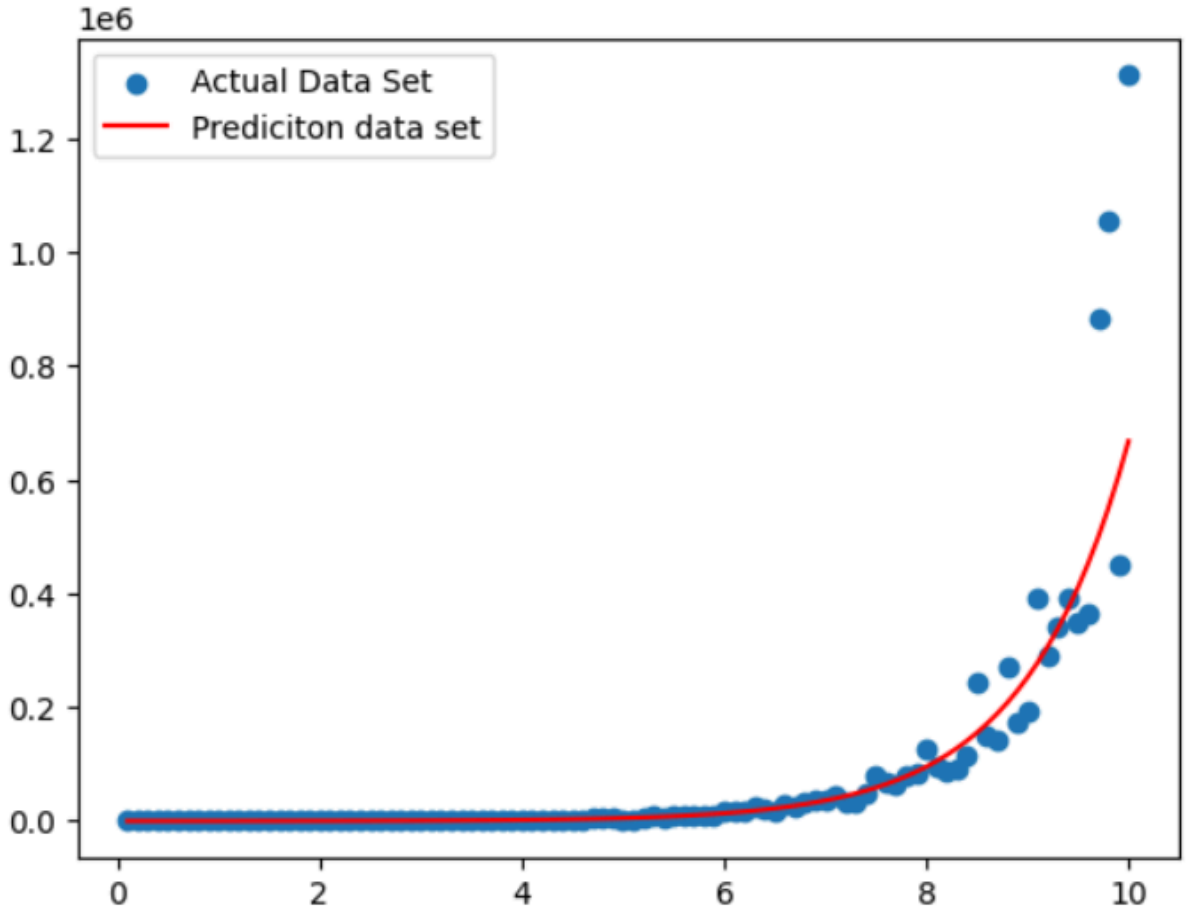
Graph of data points of data set1



First of all, on this data set we can apply standard Linear Regression. Since there are only two columns in the data set one for ‘X’ and other for ‘Y’ and if we plot these points we see blue points in the above graph. From this we get an intuition that we can directly apply standard Linear regression. Now after applying Linear regression and calculating Mean Square Error, Root Mean Square Error, Mean absolute Error, R Square our intuition is right as these values are less. If we particularly take the R Square value that is 0.95795719, we know that if the model's R Square value is close to 1 then it means the model is good for that data set which we can see. We can also take MSE, MAE, or RMSE values also, which is the summation of the difference between y and y_hat where y_hat is predicated y by our model. These all values are not large and are in a range (0, 2.5).

DataSet2:-

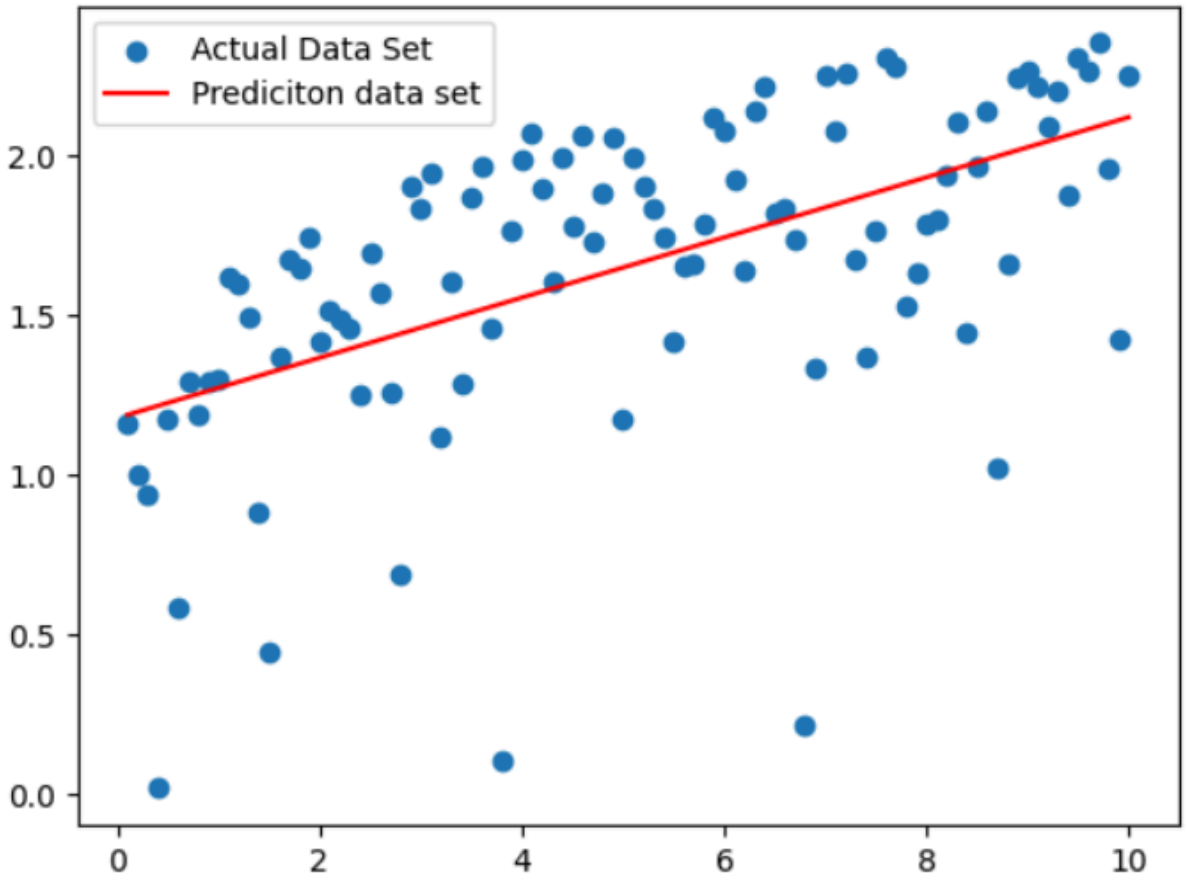
Graph of data points of data set3



For this data set we first have to do nonlinear transformation, so that we can apply standard linear regression. As we can see above, the graph is of e^x so we do transformation and then apply linear regression. Now if we see R square's value 0.7882858000815468 and this value is near to 1 so we can say that for this data set our model is good. If we see other error's value like MSE, MAE and RMSE which is 8892037637.213293, 25883.636524430738, 94297.60143934359 respectively and these values are large so from these three error's value we can say that our model is not too much good but yes it good as R square and the graph that we see above tells that there are three to four values which are too large. Due to which MSE, MAE, RMSE values are larger.

Dataset3:-

Graph of data points of data set3



Now for this data set Linear regression is not a good model or we can say that Linear regression is not applicable for this. From the above graph and then if we see R square's value that is 0.31369732, this value is near to 0 not to 1 so we can say that our model is bad for this dataset. Even though MSE, MAE, RMSE values 0.16173044, 0.29467793, 0.40215723 respectively are very less or near to Zero but when we see the graph we can clearly see that Linear regression is not a good model for this.

DataSet4:-

Since this data set has more than 3 columns, creating a graph was not possible for me. Since there are multiple columns in the data set from there we get an intuition that this might be of multiple linear regression. So after applying multiple regression and seeing R square value 0.9579571905586357 and this value is almost 1 and we know that R square's value near to 1 means for that data set our model is good. Now we also check other error values MSE, MAE, and RMSE which are 2.0785254017773265, 1.2805559784291458, 1.4417091945941547

respectively. These values are also not large enough so from all these things we can say that for this data set was Multiple Linear Regression.