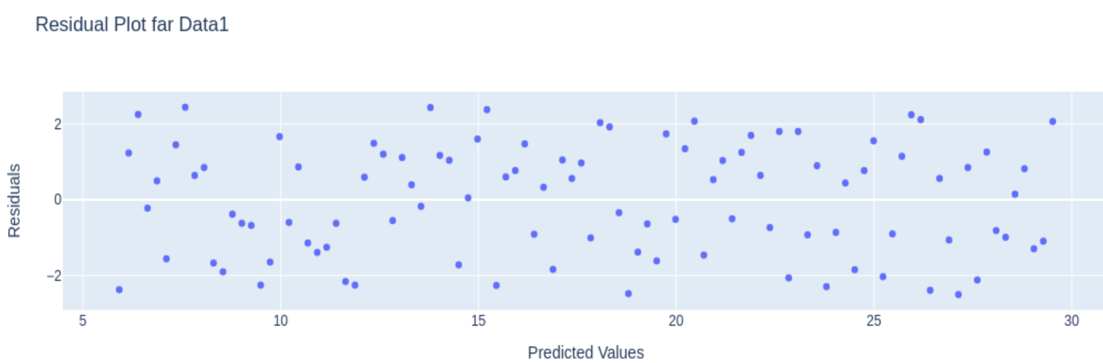
Linear Regression Analysis Report

Suraj Prajapati

Table 1 (Data 1)

Dataset	MAE	MSE	RMSE	R^2
Simple Linear Regression	1.280555978429146	2.0785254017773265	1.4417091945941547	0.9579571905586358
Gradient Descent	1.284229373056244	2.1150369286293262	1.4543166534937728	0.9572186635410976



In the above table, the different errors for Data1 have been shown. The error calculated through both algorithms , inverse matrix and using gradient descent has shown the almost same error. The R^2 value for both is close to 1. This justified that the model is the best fit model.
In the below Residual Plot the residual point(true value - predicted value) is randomly and nearly uniformly distributed along the horizontal line. Since this dataset directly gives **good results with the usual Linear Regression algorithm.** So our regression model is appropriate for the given dataset. Here our regression model is linear. You can see it from the figure below.

[Best Hyperplane Y = 5.68078713+ 2.38406007*X]

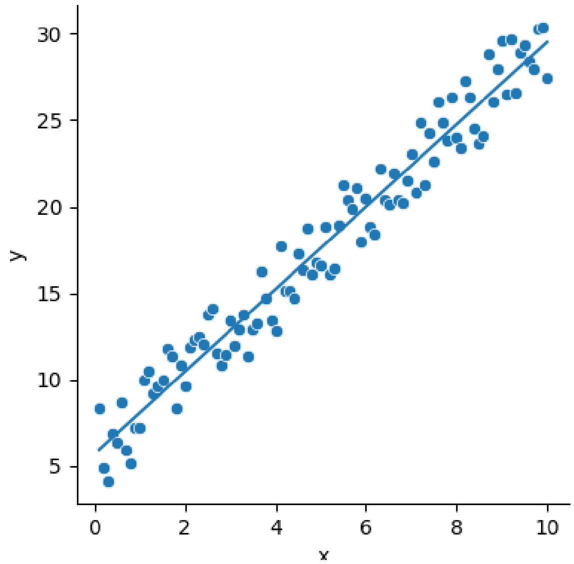
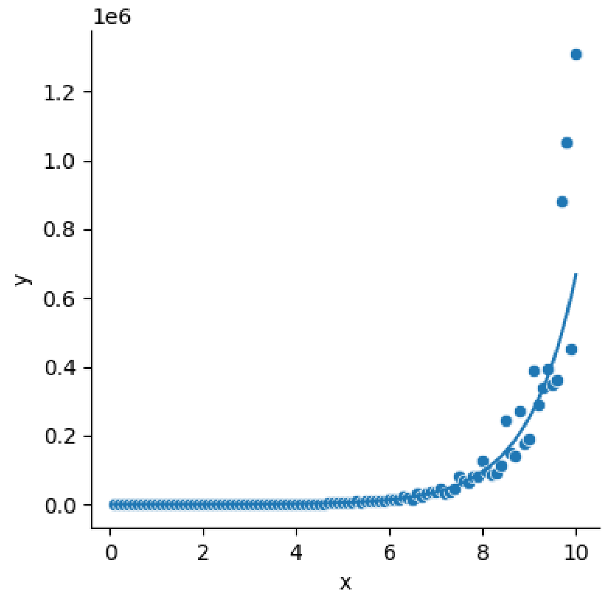


Table 2 (Data 2)

	MAE	MSE	RMSE	R^2
Simple Linear Regression	99929.7832971416	27577785853.164074	166065.6070749271	0.34338909644390125
LR using Gradient Descent	99929.76847886939	27577785853.168034	166065.607074939	0.3433890964438069
Using Logarithmic Regression	25883.636524430727	8892037637.21335	94297.60143934388	0.7882858000815314



Here the MAE, MSE and RMSE have calculated the large error and also the R^2 value is very far from around the value of 1 in case of Linear Regression. So the regression model evaluated through simple regression is not going to fit the model. Since the dataset is similar to the exponential function. So we will attempt the **non-linear transformations**. For that we will take a log of actual y value in the dataset (log(y)), then calculate the regression model to evaluate the weights. Then we will evaluate the predicted value for the above model.

On evaluating this model this gives the very reasonable MSE,MAS and RMSE but the good value of R^2 suggests a good regression model. But the residual plot of this dataset is not randomly dispersed around the horizontal line. It seemed to follow some continuous pattern. So the model for non-linear transformation of the input features would not be correct.

Predicted Hyperplane Y = exp(2.01637719 + 0.15128513*X)

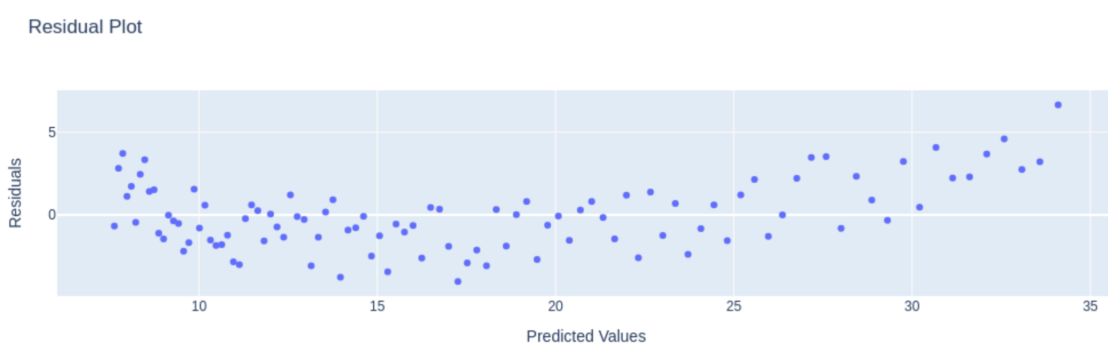


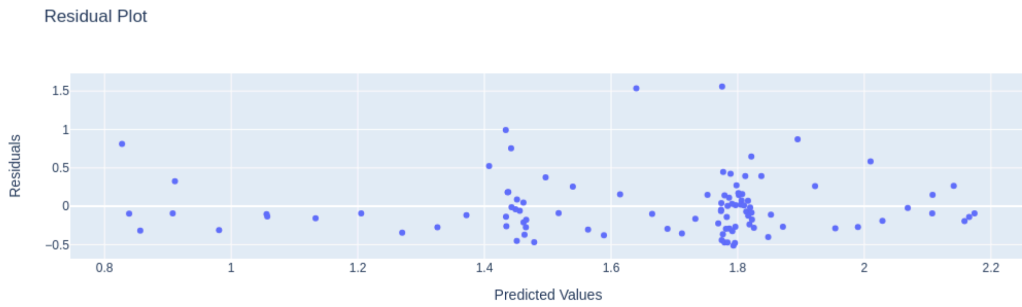
Table 3 (Data 3)

Dataset	MAE	MSE	RMSE	R^2
Linear Regression	0.29467793301310363	0.16173044143088552	0.4021572347116057	0.3136973226728079
Gradient Descent of LR	0.3076562835325742	0.16562966235089135	0.4069762429809526	0.2971509895693162
Polynomial Regression	0.2728495296917847	0.14466752744736108	0.3803518469093598	0.3861037505925682

This dataset using simple linear regression gives very bad results with R^2 value about 0.3 order. **standard Linear Regression algo is not suitable here**. Since the given data is widely dispersed over the dataset. This gives the intuition of applying polynomial linear regression so I chose different polynomials with different degrees and tried to fit it into the regression model. But I failed to generate a good value of R^2.

Predicted Regression model

Y = 1.30985339e+00 - 3.17013559e+00*X1 + 6.85787138e+00*X2 - 5.79520052e+00*X3 + 2.57277384e+00*X4 - 6.63986604e-01*X5 + 1.03269443e-01*X6 - 9.55148561e-03*X7 + 4.84264483e-04*X8 - 1.03684602e-05*X9



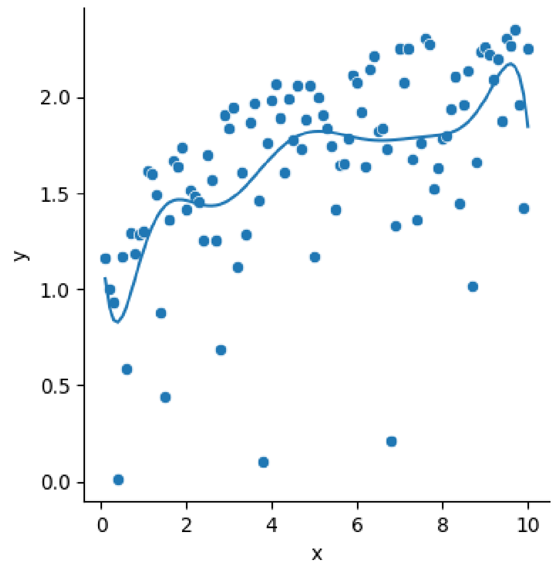


Table 4 (Data 4)

Dataset	MAE	MSE	RMSE	R^2
Linear Regression	5.155505630377769	34.62048082924355	5.8839171331047435	0.9841749058943147
Gradient Descent	5.1600499353841425	34.62989717535018	5.884717255344574	0.9841706016628395

In this dataset, the simple linear regression directly gives good output with R^2 value very close to 1. Also its residual plot has been randomly dispersed round the horizontal line.

Best fit Hyperplane y = 6.13243763*X1+ 2.39226554*X2 + 7.74681038X3 + 13.23947782