Report

- **Pankaj Yadav**

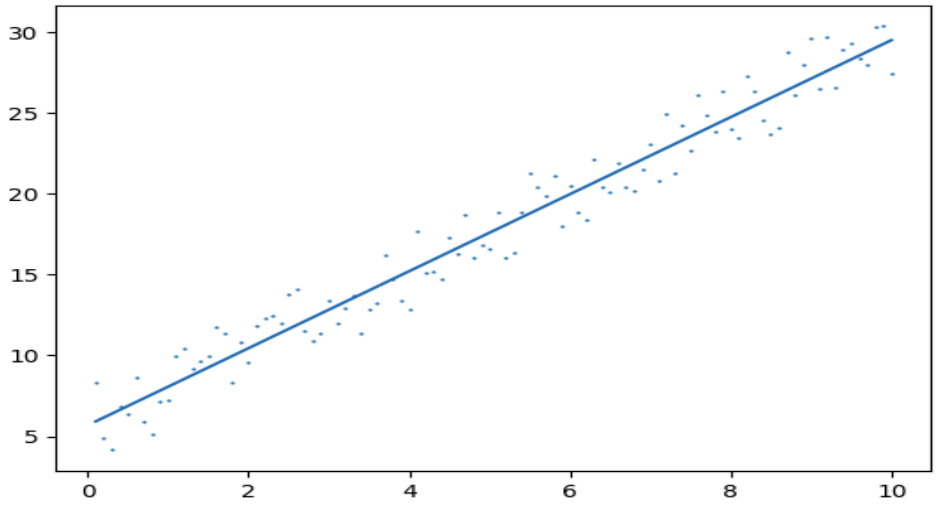
Purpose : The purpose of this report is to explain the proper reasoning about categorising datasets into different forms of Linear Regression. Here I have four datasets in which I have to find the best fit hyperplane for the four synthetic datasets.

Analysis of Datasets Using Matrix Inverse:

Datasets	Features	Features Coefficient	R Square	Remark
Dataset1	x0	5.6807871	0.9579	It follows the simple LR.
	x	2.3840600		
Dataset2	x0	3.6821226	0.78906	It requires non-linear transformation of input features.
	x	0.9729974		
DataSet3	x	-	-	Standard LR is not suitable.
Dataset4	x0	3.59	0.9775	It follows the multiple LR.
	x1	-4.21240322		
	x2	8.5411979		
	x3	12.7422045		

Hyperplanes :

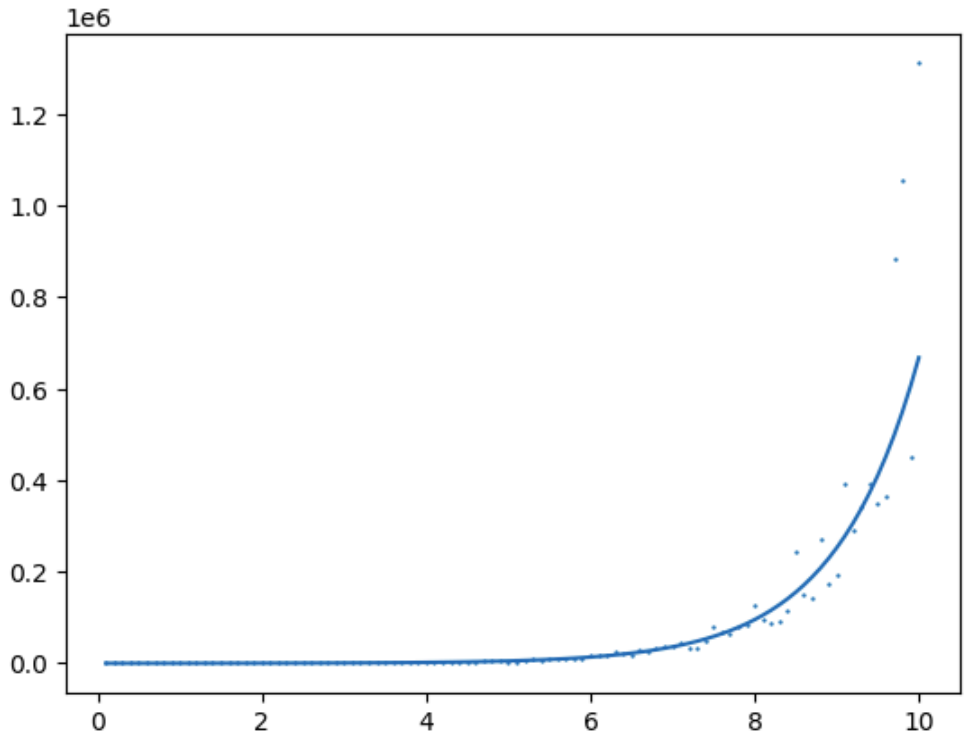
For Dataset1 -



I have predicted for the dataset1 and found that it fits the data points as a straight line (above image).

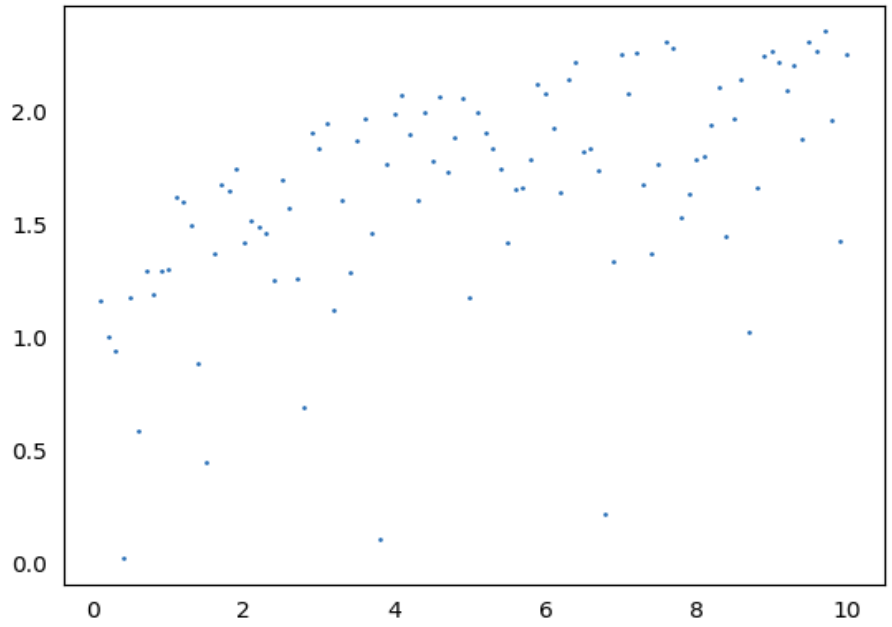
For Dataset2 -

In dataset2, I have predicted the curve and found that it fits the data points as an exponential curve. Before choosing the exponential function I have tried five to six polynomial functions but it was not fitting properly, I chose the exponential function and found the more suitable r square .



For Dataset3 -

NO, Hyperplanes exist of Linear Regression in this dataset because we can't apply LR. Because here we are not able to see any straight line or curve to apply LR.



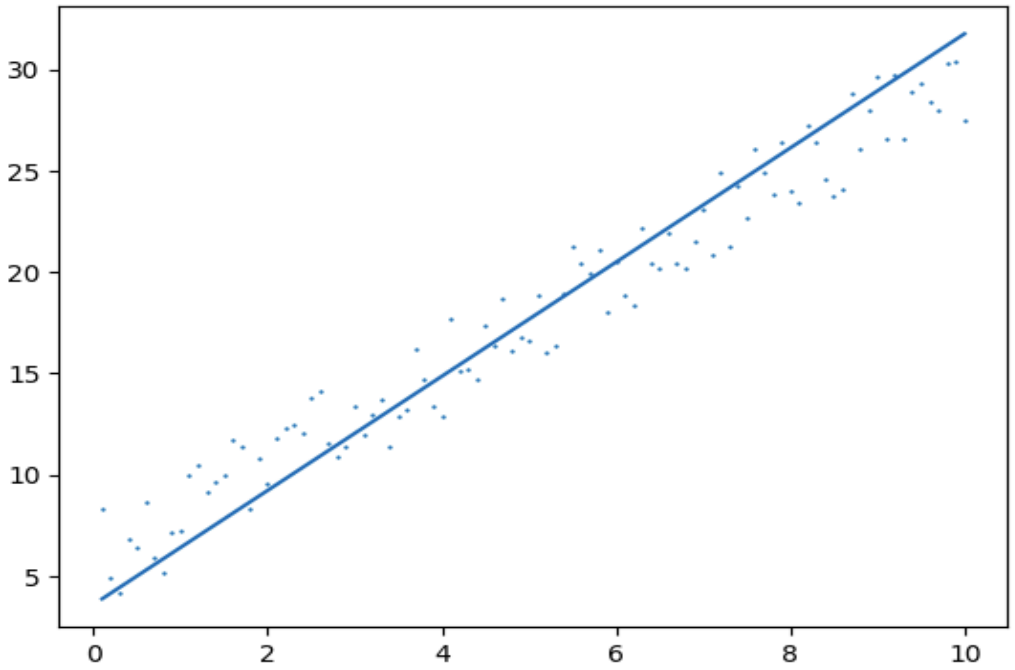
For DataSet4 -
Here Hyperplanes exist but we can't see them visibly. I have used the concept of multiple linear regression and found, r square 0.9775 which is very close to the 1. So it follows LR.

Analysis of Datasets using Gradient Descent -

Datasets	Features	Features Coefficient	Learning Rate	Epoch	R Square	Remark
Dataset 1	x0	3.59	0.00001	3330	0.92635	It follows the simple LR.
	x	2.8160493				
Dataset 2	x0	0.8	0.00001	3330	0.796816	It requires non-linear transformation of input features.
	x	1.3266423				
Dataset 3	x	-	-		-	Standard LR is not suitable.
Dataset 4	x0	3.59	0.00001	239	0.97195	It follows the multiple LR.
	x1	5.3617501				
	x2	5.8142769				
	x3	6.2744345				

Here, I have minimised the squared error using GD. Here I use BGD where I change or update the value of weight after calculating the gradient for whole datasets. And keeping the LR at 10^-4 for each datasets and iterating over corresponding epoch and found the coefficient of features.

For Dataset1 -



For Dataset2 -