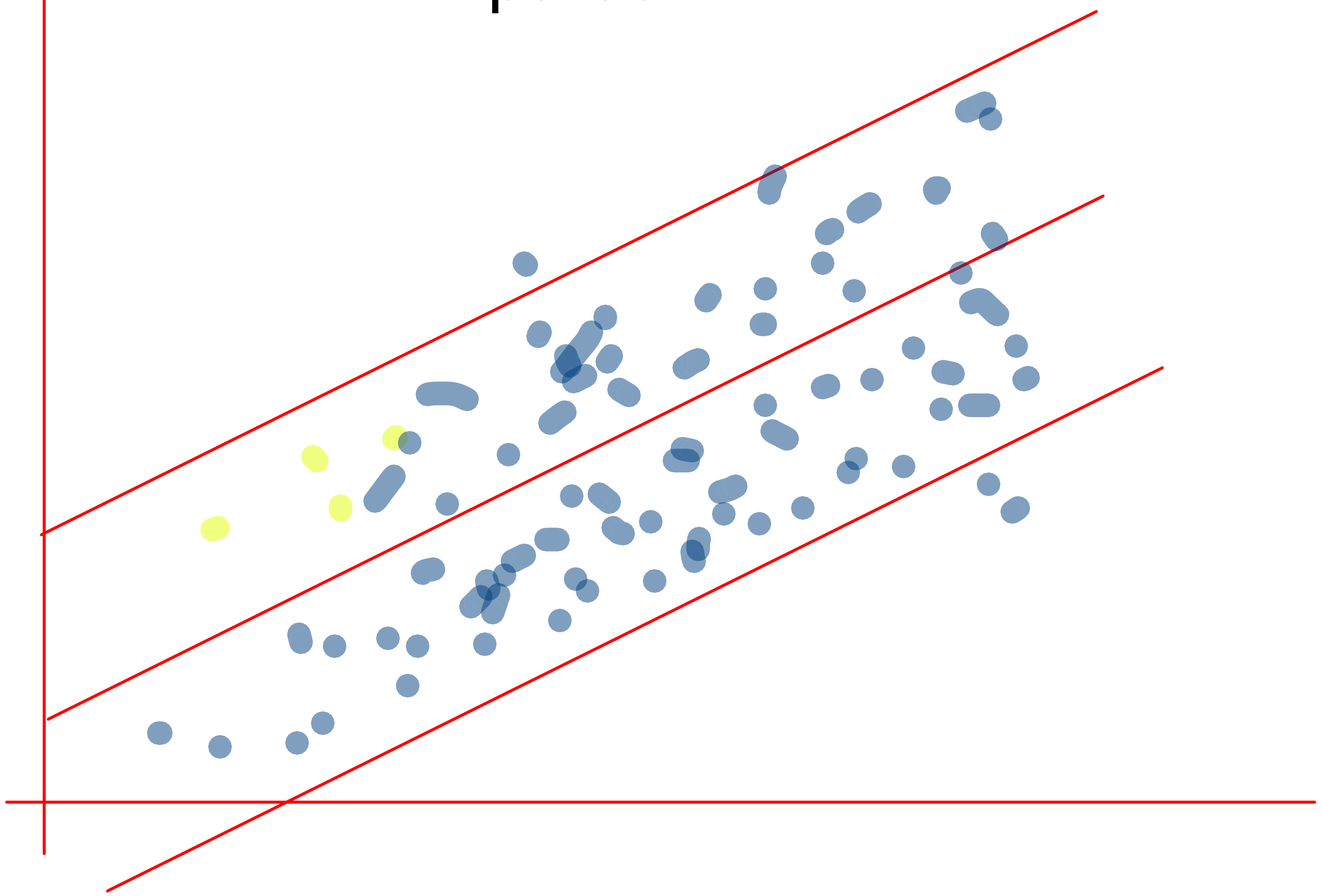
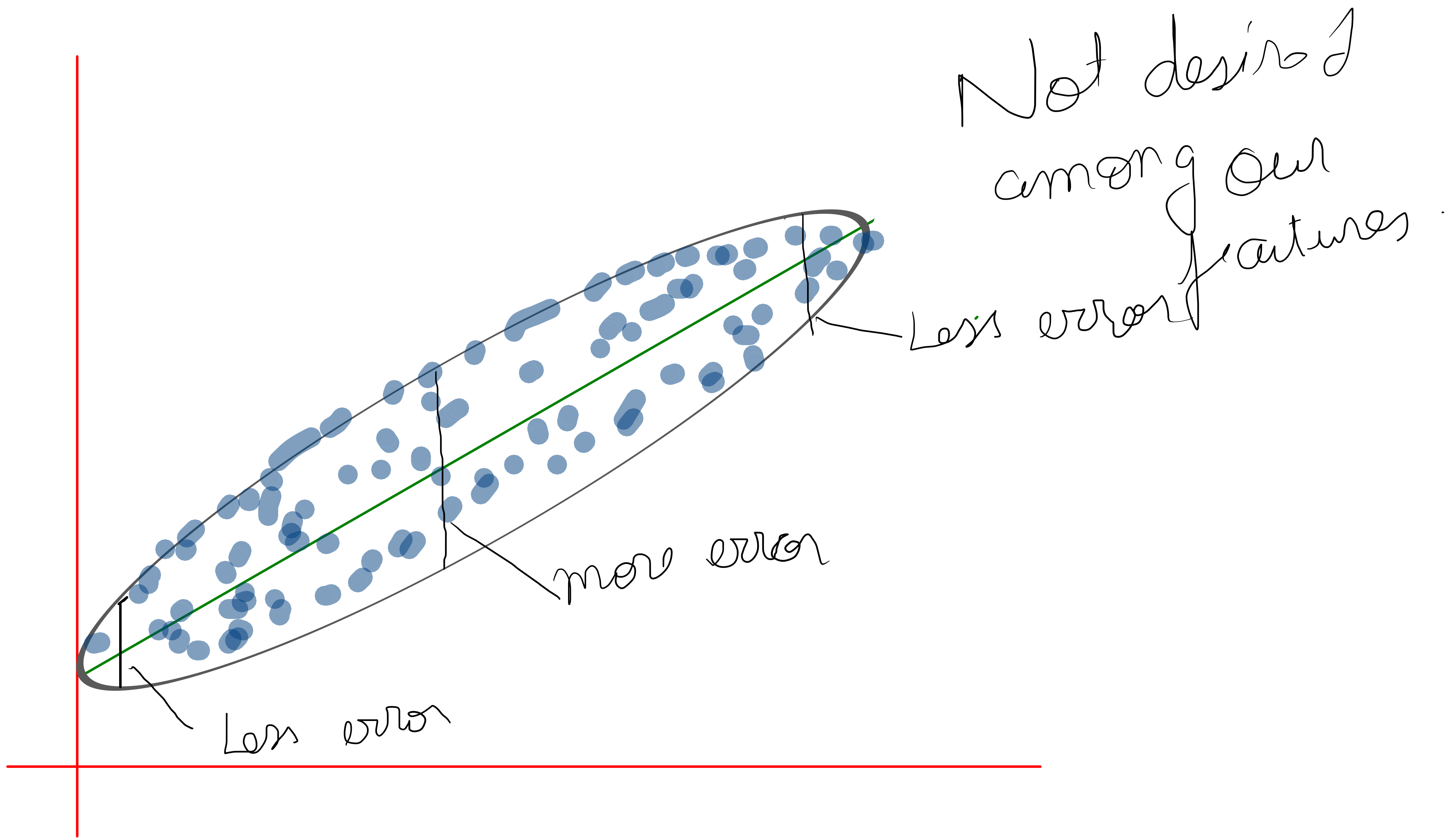
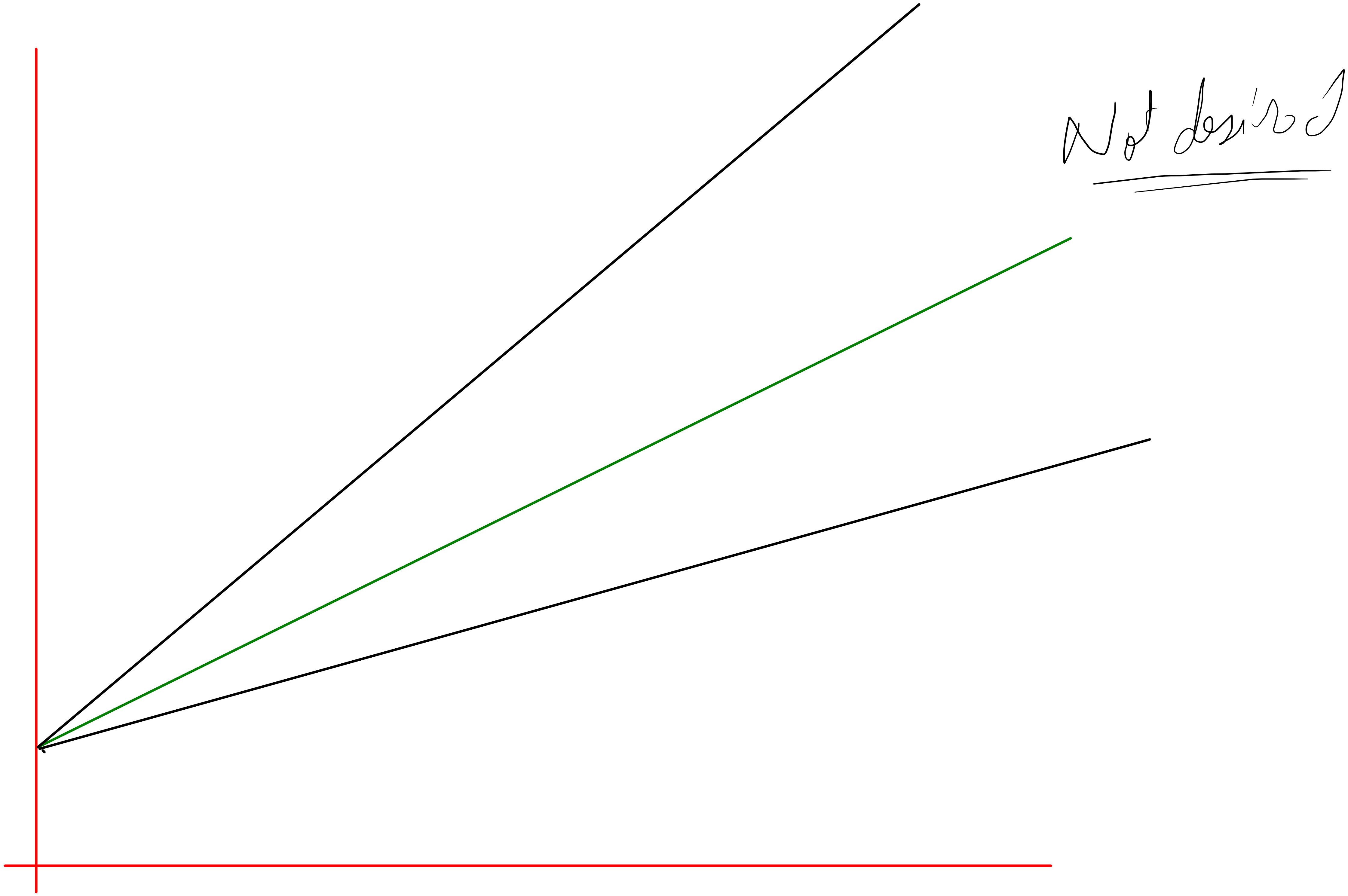
Homocedacity : That there is no pateern

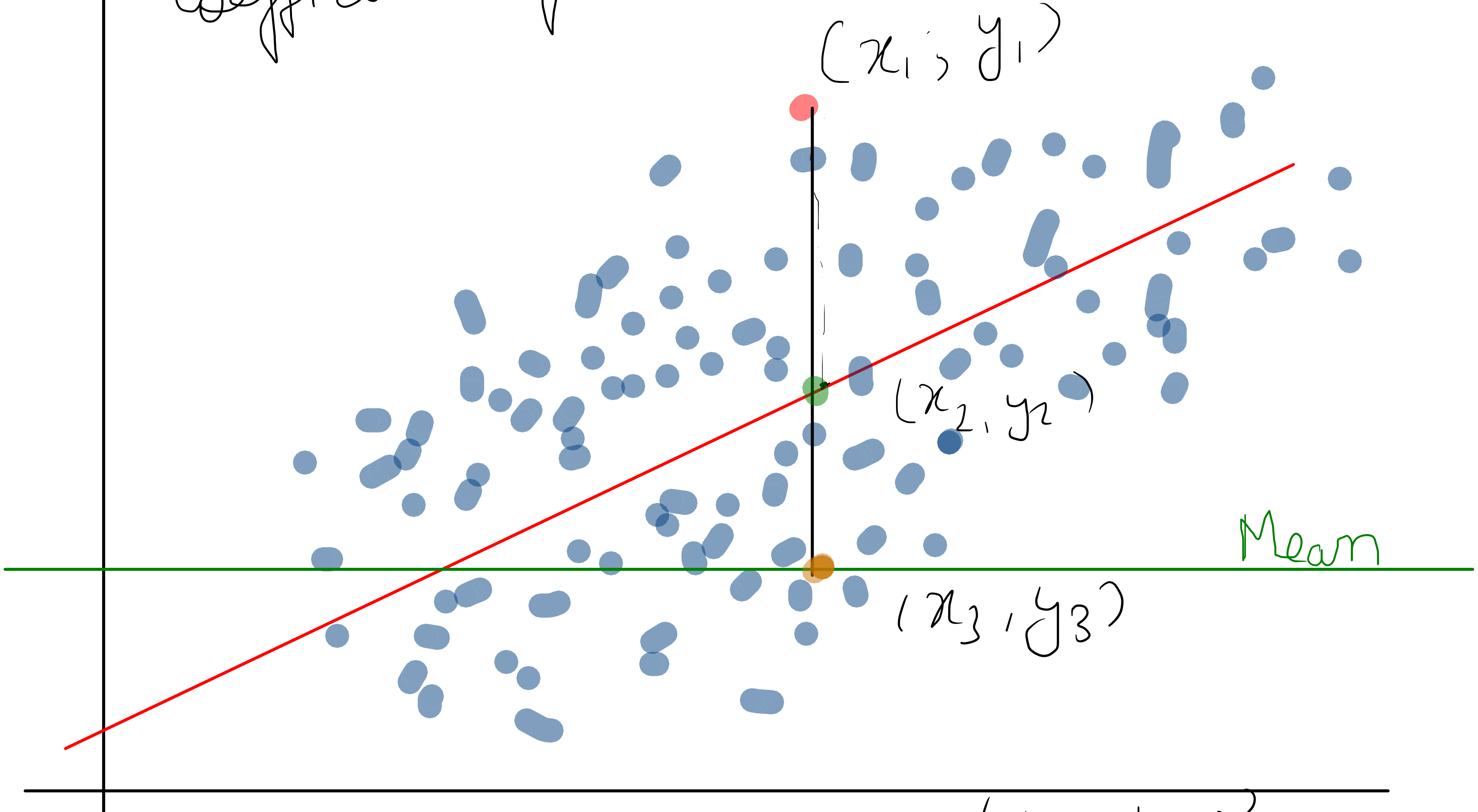




Not desired among our features.
Less error
more error
Less error


Not desired

Coefficient of Determinant:



Sum Square error = $\sum (y_1 - y_2)^2$

Sum Square residual (SSR) = $\sum (y_2 - y_3)^2$

Sum Square

Coefficient of Determinant (R^2) / R square

$$= \frac{SSR}{SST} = \frac{SSR}{SSE + SSR}$$

I conclude that = higher R^2 better is my model.

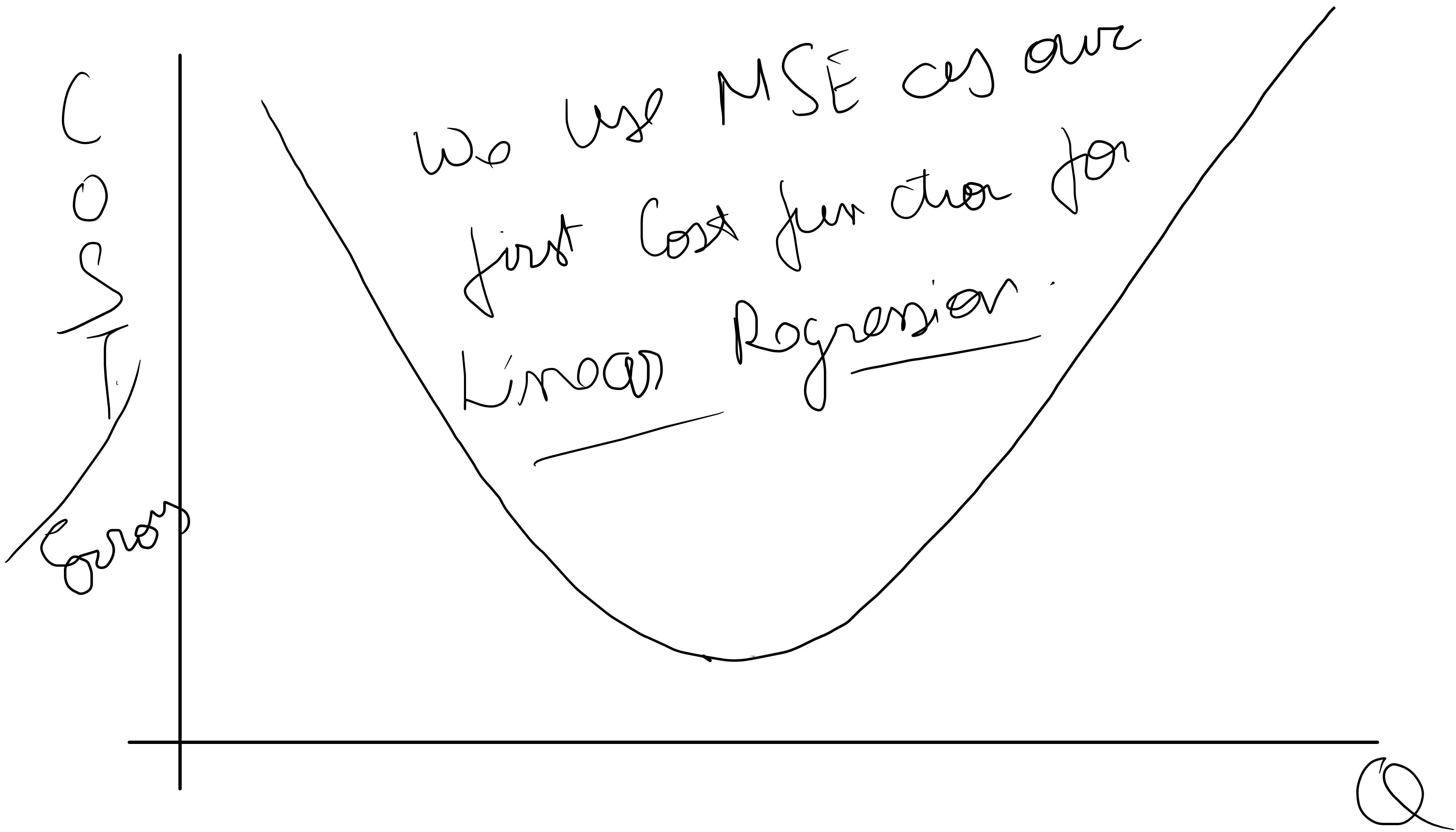
MAPE = Mean absolute percentage error

$$= \sum_{i=1}^{n} \left| \frac{\cancel{\Sigma}(\hat{y} - y_i)}{y_i} \times 100 \right|$$

$$RMSE = \sqrt{MSE}$$

↳ Use this for showing data/result to management/marketing.

The function / value which minimizes the MSE will also minimize RMSE.



$y(\text{price}) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n + C$
$+ \varepsilon(\text{error})$

For above equation there may be some non-significant features included in our model.

Q) How to we deal with this problem of insignificant features?

Ans Adjusted $R^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$

R^2 = coefficient of determinant   n = number of instances / number of rows

k = no. of features / columns.

Adj R^2 = punishes our model for adding insignificant terms to our model.

Adjusted R^2 will always be $\leq R^2$ if we are using all significant features than Adj $R^2 = R^2$

Mathematically Adj R^2 can be -ve but in that case our R^2 will be very low. hence there is some problem with our mode drastically he we will ditch this model & make a new one.

What about MSE for the same problem.

1) Ridge regression

2) Lasso "

3) Elastic Net.

Cost function $(J(\theta))$ = MSE + Regularization Term.

Ridge

$$= MSE + \boxed{\frac{\alpha}{2} \sum \theta_i^2}$$

manually set this parameter.

$\sum \theta_1^2 = (.9)^2 = (.81) \downarrow$

$\sum \theta_1^2 = (1.2)^2 = (1.44) \uparrow$

$10 \Rightarrow \underline{10.81}$ or $\underline{11.4}$

<u>154</u>

$$\text{Lasso} = MSE(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

Lasso helps us in finding insignificant features. as it reduces the weight of insignificant features to close to 0 / zero.

$$\hat{y} = h_\theta(x) = \theta^T \cdot x$$

$$\hat{\theta} = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

for a data set of 1000 rows & 1 column

it will take 100 secs to calculate

result.