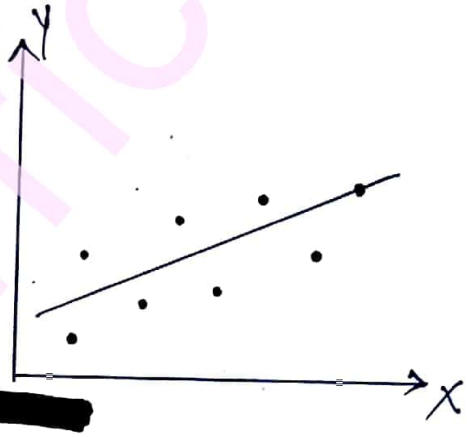
Simple Linear Regression (Theory)

In statistics, linear regression is a linear approach to model the relationship between a responsive variable (say y) and one or more explanatory variable (say x).

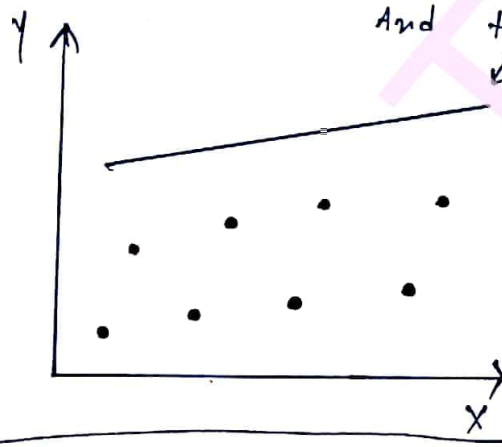
The case of only one explanatory variable is called Simple Linear Regression.

In Machine Learning it comes under Supervised Learning.

Simple Linear Regression →

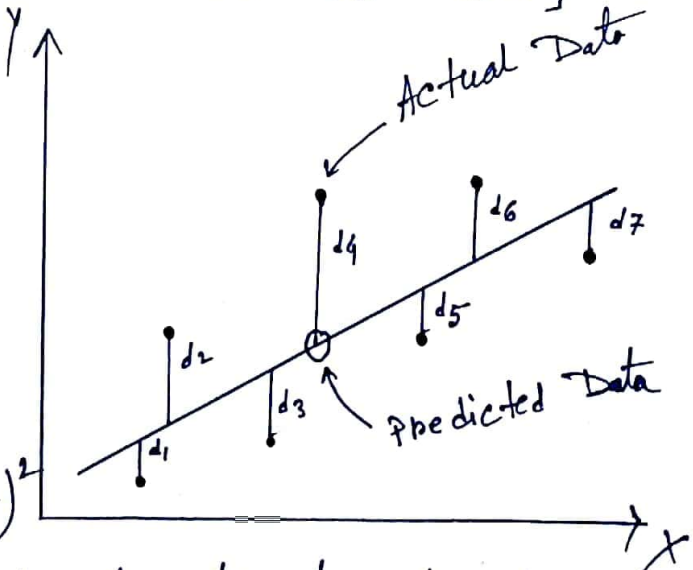


Suppose, my model has predicted a line like this → [We have to find the best fitted line And this is not the best fitted line.]



So, the least square Method is

= SUM (Actual point - Predicted point)^2



, Suppose, the distances are, $d_1, d_2, d_3, d_4, d_5, d_6, d_7$

$$\Rightarrow D = (d_1)^2 + (d_2)^2 + (d_3)^2 + (d_4)^2 + (d_5)^2 + (d_6)^2 + (d_7)^2$$

[Tec4Tric]

Best linear model [redacted] will have the smallest value of D.

Now, what is the equation of straight line →

$$y = a + bx \qquad [\hat{y} = b_0 + b_1 x]$$

$\hat{y}$ = predicted, a = intercept, b = slope, x = predictor

[x = Explanatory variable / Independent
y = Responsive variable / Dependent]

We have to find a & b.

$$a = \bar{y} - b\bar{x} \qquad b = R \cdot \frac{S_y}{S_x}$$

[R = Pearson correlation coefficient,
S_y = Standard Deviation of y,
S_x = Standard Deviation of x]

$$R = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

$$S_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n-1}}$$

$$S_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

[n is the number of Observations]

[Tec4Tric]