Linear Regression

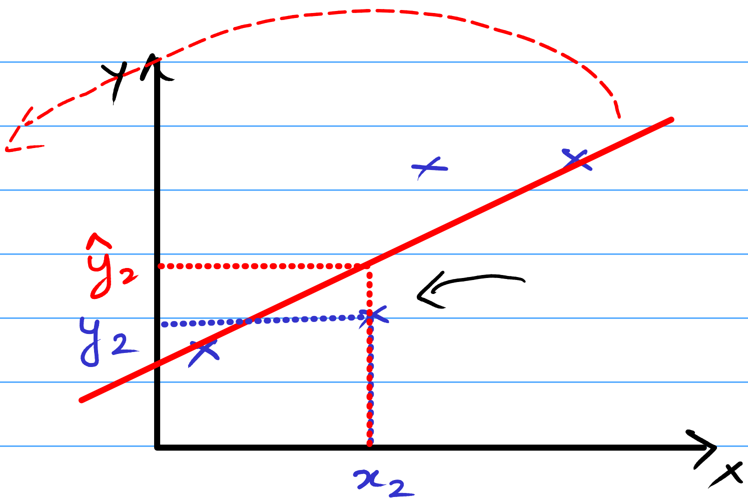
Data

Discrete	Continuous
→ finite Data, and hopefully not very large	→ infinite Data
→ You can drill down to a particular data point.	→ You cannot drill down to a particular data point, instead you get a range.

↓

Regression

$\hat{y} = mx + c$
(Model)
in Pg 2



Actual Dataset

X	Y
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4

X → independent Variable

y → Dependent (on x) variable.

if Input (I/P) : x_2

then Actual O/P : y_2

Predicted O/P : $\hat{y}_2$

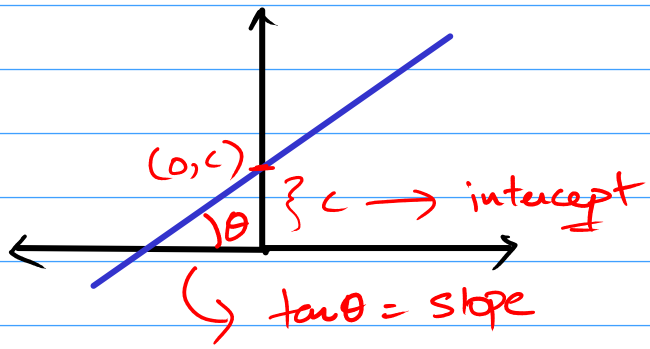
Error : $(y_2 - \hat{y}_2)$

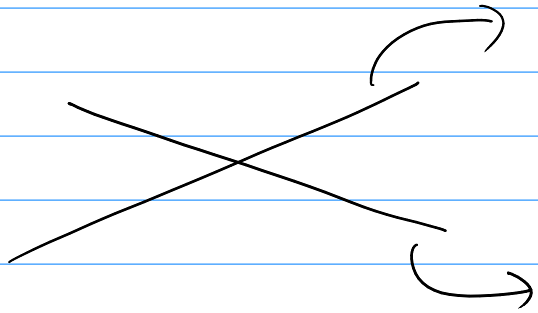
Model:

Predicted Model $\rightarrow$ $\hat{y} = mx + c$

where $m = \tan\theta$ (slope)

c = intercept

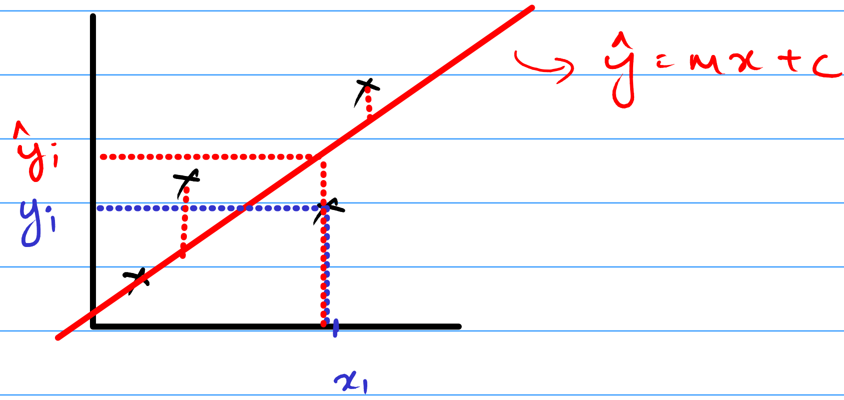




$l_1 : y = m_1 x + c_1$

$l_2 : y = m_2 x + c_2$

Errors:



$$\text{Square of Error} = (y_i - \hat{y}_i)^2$$

$$\underset{(SSE)}{\text{Sum of Squared Errors}} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Finding the appropriate value(s) of the slope(s) and the intercept such that SSE is the least. $\Rightarrow$ Best Fit Model.

Given n i/Ps & o/Ps :

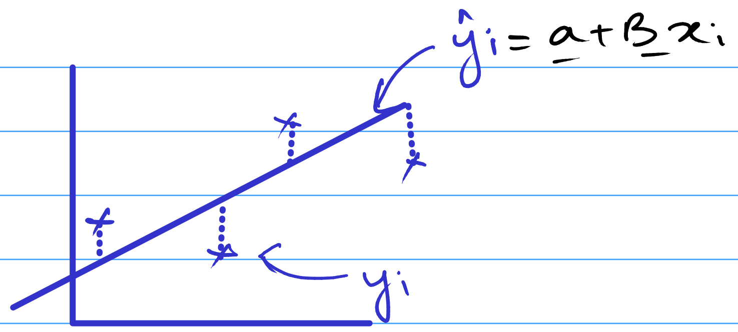
$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

Predicted Model :

$$\hat{y}_i = a + Bx_i \quad \text{— ①}$$

where a = intercept

B = Slope (coefficient)



$$J = SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

from eqⁿ ①, we get

$$J = \sum_{i=1}^{n} \left(y_i - (a + B x_i) \right)^2$$

$$J = \sum_{i=1}^{n} (y_i - a - B x_i)^2 \quad \text{--- ②}$$

To find the least Cost Function wrt a & B, we will partially differentiate J wrt a and B.

$$\frac{\partial J}{\partial a} = 0 \qquad\qquad \frac{\partial J}{\partial B} = 0$$

Finding a :

$$\frac{\partial}{\partial a} \left(\sum_{i=1}^{n} (y_i - a - B x_i)^2 \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} -2 (y_i - a - B x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - B \sum_{i=1}^{n} x_i = 0$$

$$\overbrace{a + a + a + \cdots + a}^{n}$$

$$= na$$

$$\Rightarrow \sum_{i=1}^{n} y_i - na - B\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \quad a = \sum_{i=1}^{n} \frac{y_i}{n} - B\sum_{i=1}^{n} \frac{x_i}{n}$$

$$\frac{y_1+y_2+\cdots+y_n}{n} \qquad \frac{x_1+x_2+\cdots+x_n}{n}$$

$$\overline{y} \qquad \overline{x}$$

$$\boxed{a = \overline{y} - B\overline{x}}$$

Finding B:

$$\frac{\partial}{\partial B} J = 0 \Rightarrow \frac{\partial \sum_{i=1}^{n}(y_i - a - Bx_i)^2}{\partial B} = 0$$

$$\Rightarrow \sum_{i=1}^{n} -2x_i\,(y_i - a - Bx_i) = 0$$

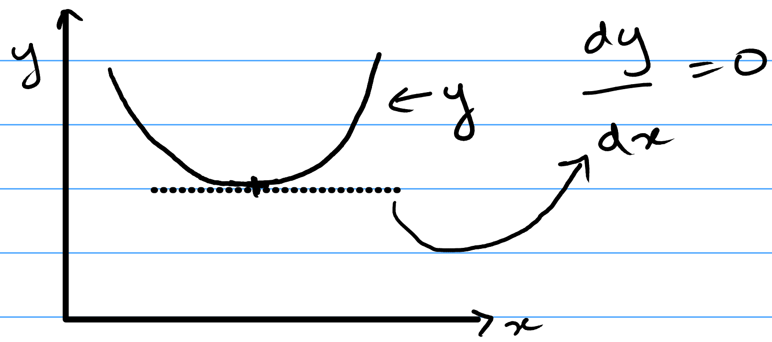
$$\Rightarrow \sum_{i=1}^{n} x_i\,(y_i - a - Bx_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - a x_i - Bx_i^2) = 0$$

Now substituting the value of a that we got moments ago, & doing the app. algebra, we get,

$$\boxed{B = \frac{\sum_{i=1}^{n} (x_i y_i - \bar{y} x_i)}{\sum_{i=1}^{n} (x_i^2 - \bar{x} x_i)}}$$

$\hat{y}_1 = a + B x_i$ (Pluggin a & B here)



$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad ✓$$

$(-5)^2 \to 25$

$(5)^2 \to 25$

$$\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad ✗$$

$|-5| = 5$

$|5| = 5$