① ML

Unit-2

⇒ Linier Regretion → supervised Learning base on

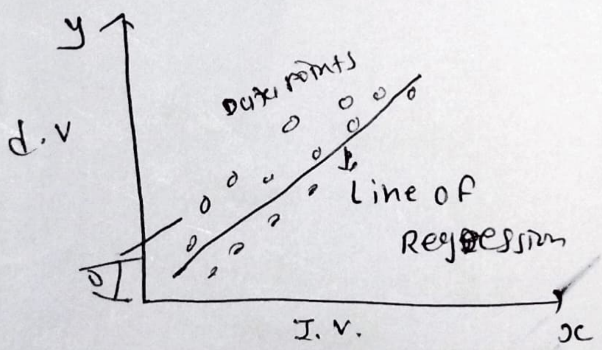
- Simple linear Regretion:-
- Multiple Linear Regretion

→ It is a statistical method that is used for predictive analysis.

→ L.R makes Predictions for continuous / real or numeric variables such as sales, salary, age, product, price etc.

→ L.R Algo show a linear relationship between a dependent variable and one or more independen variable its called L.R.

→



$y = a_0 + a_1 x + \varepsilon$

$\boxed{y = mx + c}$

☆ types of L.R

- simple:- If a single independent variable is used to predict the value of numerical ~~predict~~ dependent ~~the value~~ variable, then such a L.R is called simple

- $y = \alpha_0 + \alpha_1 x_1 + \varepsilon$

α_i = Reg. coefficent

x_i = Independent

y = dependent

* Multi L.R

→ If more than one Independent Variable is used to Predict the value of numerical dependent variable

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m$$

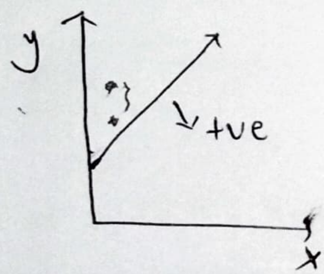
* L.R Line

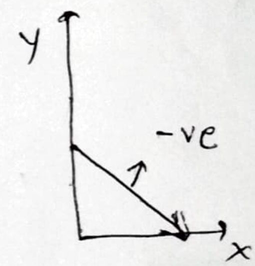
A linear line showing the relationship between the dependent and independent variable is called r. Line.

two type relationship :-

- Positive L.R line
- Negative L.R line

Positive⇒





- dependent variable increases on y axis
- independent on x

- dependent decreases on y
- independen increases on x

* eq of L.R

Line = $\boxed{y = mx + c}$ | $y = ax + b$

m-slop

$$a = \left(\frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sum (x - \bar{x})^2} \right)$$

$$a = \frac{n(\sum x \cdot y) - (\sum x)(\sum y)}{n \sum (x^2) - (\sum x)^2}$$

$$b = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

or

$$b = \frac{1}{n}\left[\sum (y) - \sum (x) \cdot a\right]$$

3

* ~~MAE,~~
* MAE (mean absolute error)

$$= \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

→ $\bar{x} = \frac{\Sigma x}{n}$ $\bar{y} = \frac{\Sigma(y)}{n}$ (mean)

→

* MSE (mean squared error)

$$= \frac{1}{N}\sum_{i=1}^{N}\left[(y_i - \hat{y})^2\right]$$

$\hat{y}$ (predicted)

* RMSE (Root mean squared error)

$$= \sqrt{MSE}$$

$$= \sqrt{\frac{1}{N}(\Sigma(y-\hat{y})^2}$$

* R^2 (coefficent determined or g

$R^2 =$

$$\frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2} = R^2$$

$R^2 =$

* Example

x	y	x·y	x^2	p ($\hat{y}$)	Error
1	3	3	1	2.8	0.2
2	4	8	4	4.1	0.1
3	5	15	9	5.4	0.4
4	7	28	16	6.7	0.3
$\Sigma x =$ 10	Σy 19	54	30		1

$n = 4$ 0.4

$y = ax + b$

$a =$

$$a = \frac{n(\Sigma x.y) - \Sigma(x).(\Sigma(y))}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{4(54) - (10)(19)}{4(30) - (10)^2} = \frac{216 - 190}{120 - 100}$$

$$= \frac{26}{20} = \frac{13}{10} \boxed{1.3}$$

$$b = \frac{1}{n}\left[\Sigma(y) - \Sigma(x).a\right]$$

$$= \frac{1}{4}\left[(19) - (10)(1.3)\right]$$

$$= \frac{1}{4}\left[19 - 13\right]$$

$$= 6/4$$

$$= 3/2$$

$$\boxed{= 1.5}$$

$\underline{y = ax + b}$

$\boxed{y = 1.3x + 1.5}$ Line eq or Linear eq

** Prediction x or y

x=1

$y = (1.3)(1) + 1.5$

$y = 2.8$

x=2

$y = (1.3)(2) + 1.5$

$= 4.1$

x=3 $y = (1.3)(3) + 1.5$

$= 5.4$

x=4

$y = (1.3)(4) + 1.5$

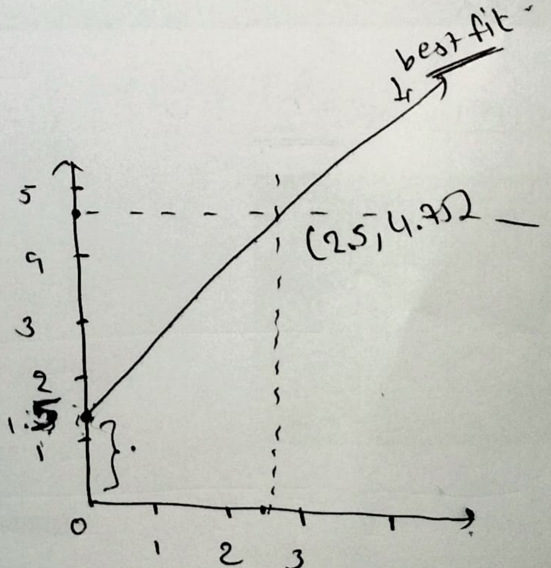
$= 6.7$

* → Error ⇒ x − P = x

→ best fit of given data

$(\bar{x}, \bar{y})$. $\bar{x} = \frac{10}{4} = 2.5$

$\bar{y} = \frac{19}{4} = 4.75$



* Stander error of estimet

$$= \sqrt{\frac{\Sigma(\hat{y}-y)^2}{(n-2)}}$$

$(y-\hat{y})^2$
0.04
0.01
0.16
0.09

0.3

$$= \sqrt{\frac{(0.3)}{(4-2)}}$$

$$= \sqrt{0.0275}$$

$$= \boxed{\sqrt{0.15}} < 1$$

$$\sqrt{\frac{0.3}{(4-2)}}$$

$$\sqrt{\frac{0.3}{2}}$$

$$= \sqrt{0.15}$$

* MAE (mean Absulute Error)

$$= \frac{1}{4}[1]$$

$$= \frac{1}{4}$$

$$= 0.25$$

* MSE $= \frac{1}{4}[0.3]$

$$= 0.075$$

* RMSE $= \sqrt{0.075}$

$$= 0.2738$$

$R^2 = \Sigma(y_i - \bar{y})$

-1.75
-0.75
+0.25
2.25

0

$$\frac{\Sigma(\hat{y}_i - \bar{y}_i)^2}{\Sigma(y-\bar{y})^2}$$

Roll.no	x Maths	y ML	x.y	x^2	$\hat{y}$	$(y-\hat{y})$	$(y-\bar{y})$	$(y-\bar{y})^2$
1	95	85	8075	9025	82.9	2.1	8	64
2	85	95	8075	7225	79.46	15.54	18	324
3	80	70	5600	6400	77.77	-7.7	-7	49
4	70	65	4550	4900	74.17	-4.17	-12	144
5	60	70	4200	4200	70.64	-0.64	-7	49
	390	385	30,500	31,750		5.13	0	630

$\bar{y} = \frac{385}{5} = 77$

$\bar{x} = \frac{390}{5} = 78$

$$a = \frac{5(30500) - (390)(385)}{5(31750) - (390)^2}$$

$$= \frac{152500 - 150150}{158750 - 152100}$$

$$= \frac{2350}{6650} = \frac{235}{665} = 0.353$$

$$b = \frac{1}{5}\left[385 - (390)(0.353)\right]$$

$$= \frac{1}{5}\left[385 - 137.67\right]$$

$$= \frac{247.33}{5} = 49.466$$

$y = ax + b$

$$\boxed{y = 0.353(x) + 49.46}$$

$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
5.9	34.81	4.41
2.46	6.05	241.49
0.7	0.49	59.29
-2.83	8	17.38
-6.36	40.4	0.4096
	89.75	$(y-\hat{y})^2 = 322.9796$

$$R^2 = 1 - \frac{\Sigma(y_i-\hat{y})^2}{\Sigma(y_i-\bar{y})^2}$$

$$= 1 - \frac{(322.97)}{630}$$

$$= 0.5126$$

$$\boxed{R^2 = 0.4874}$$

$$\frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2}$$

$$= \frac{89.75}{630}$$

$$\boxed{= 0.1482}$$

* MSE

$= \frac{1}{5}[322.97]$

$= 64.59$

* RMSE

$= \sqrt{64.59}$

$= 8.036$

* Stunder error estimation Residuals

$\Rightarrow \sqrt{\frac{\sum(\hat{y}-y)^2}{(5-2)}}$

$= \sqrt{\frac{89.75}{3}}$

$= \sqrt{29.91}$

$\boxed{= 5.46}$