Linear Regression →

Linear Regression is a Method to Predict dependent variable(y) based on values of independent variable(x). it Can be used for the Cases where we want to Predict some Continuous quantity.

for Example

Suppose you have to find out annual Sale of your Company.

So factors that affect the Annual Sale known as dependent variables.

like.
* Employee
* Product Sale
* loss
* Profit

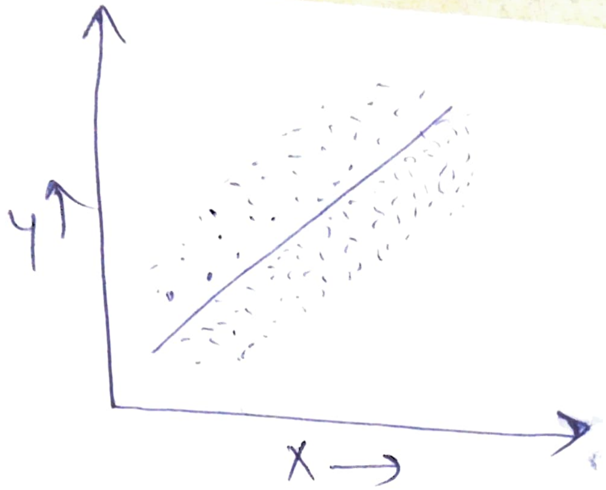
] Dependent Variables

* Annual Sale — Independent variable

Regression analysis is used for Establishing the relationship between the dependent & independent variables.

OR

When dependent variable or output. if we want in numeric or Number then we will use regression.



linear Means — When x increase y is also increase this is linear nature.

* Simple linear Regression →

if we have only one dependent Variable and one independ variable then we can say that it is Simple linear regression.

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

dependend variable

independent variable

* Multiple linear Regression →

if we have one dependent variable and More than one independent variable then we can say that it is Multiple linear Regression.

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 \cdots \alpha_n x_n$$

α_1 = Regression Coefficient

x_1 = independent variable

y = dependent variable

$$y = \underset{\alpha_1}{0.9} + \underset{\alpha_2}{1.2}\,x_1 + \underset{\alpha_3}{2}\,x_2 + \underset{\alpha_4}{4}\,x_3 + \underset{\alpha_5}{1}\,x_4$$

The value of coefficient shows that the features related to the coefficient How much important. it is important for output

it shows the importance of the features

In the above equation the most important feature is α_4 because it's value is equal to 4 that is maximum.

for Example

Student → Percentage

Student → Roll No/ID

if we have Percentage than we can decide that student is fail or Pass. So Percentage feature is More important to determine the result. So it is More Effective feature.

From Where α value Comes →

4

Whole data set to When we give training phase then learning process starts then Model automatically decide Which feature is putting More impact on the output. and How Much it is Important for the output.. according to this it decides α values.

Ex if we have only Student's ID and roll No then we can not Predict the result of the Student.. So these features are not Important for output.

Given Dataset

features
attributes
Parameter
Variables

X	Y	XY	X^2
1	3	3	1
2	4	8	4
3	5	15	9
4	7	28	16
10	19	54	30

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

→ ~~No~~ of observation
→ No of cases
→ No of values
→ No of Points
*You can say anything

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

$$\boxed{Y = bx + a}$$

$$a = \frac{(19)(30) - (10)(54)}{(4)(30) - (100)}$$

$$= \frac{570 - 540}{120 - 100}$$

$$= 30/20$$

$$= 3/2 \Rightarrow 1.5$$

$$b = \frac{(4)(54) - (10)(19)}{(4)(30) - (100)}$$

$$\Rightarrow \frac{216 - 190}{120 - 100}$$

$$= \frac{26}{20}$$

$$= \frac{13}{10} \Rightarrow 1.3$$

$$\boxed{Y = 13x + 15}$$

Alternative formula of (a)

$$a \Rightarrow \frac{\Sigma y - b(\Sigma x)}{n}$$

$$\frac{19 - (1.3)(10}{4} \Rightarrow \frac{6}{4} = \frac{3}{2} = 1.5$$

Multiple linear Regression →

Q Predict the value of y given x_1 and x_2

Subject	y	x_1	x_2
1	-3.7	3	8
2	3.5	4	5
3	2.5	5	7
4	11.5	6	3
5	5.7	2	1
Testing Example → 6	?	3	2

there are five instances

here five instances given So using these five instances we need to Model and then use that Particular Model ~~do~~ ~~need to~~ find the ~~Res~~ value of y.

$$a = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2$$

$$b_1 = \frac{(\Sigma x_2^2)(\Sigma x_1 y) - (\Sigma x_1 x_2)(\Sigma x_2 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$b_2 = \frac{(\Sigma x_1^2)(\Sigma x_2 y) - (\Sigma x_1 x_2)(\Sigma x_1 y)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

Subject	y	X_1	X_2	$X_1 X_1$	$X_2 X_2$	$X_1 X_2$	$X_1 y$	$X_2 y$
1	-3.7	3	8	9	64	24	-11.1	-29.6
2	3.5	4	5	16	25	20	14	17.5
3	2.5	5	7	25	49	35	12.5	17.5
4	11.5	6	3	36	9	18	69	34.5
5	5.7	2	1	4	1	2	11.4	5.7
Σ	19.5	20	24	90	148	99	95.8	45.6

x_2 is Not a simple x_2^2 it is the variance with respect to x_2. $\Sigma x_i^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{N}$

$i=1$

$$\Sigma x_1^2 = \Sigma x_1 x_1 - \frac{(\Sigma x_1)(\Sigma x_1)}{N} \Rightarrow 10$$

$i=2$

$$\Sigma x_2^2 = \Sigma x_2 x_2 - \frac{(\Sigma x_2)(\Sigma x_2)}{N} = 32.8$$

$$\Sigma x_1 y = \Sigma x_1 y - \frac{(\Sigma x_1)(\Sigma y)}{N} = 17.8$$

$$\Sigma x_2 y = \Sigma x_2 y - \frac{(\Sigma x_2) \cdot (\Sigma y)}{N} = -48$$

$$\Sigma x_1 x_2 = \Sigma x_1 x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{N} = 3$$

$$b_1 = \frac{32.8 * 17.8 - 3 * (-48)}{10 * 32.8 - 3 * 3} = 2.28$$

$$b_2 = \frac{10 * (-48) - 3 * 17.8}{10 * 32.8 - 3 * 3} = -1.67$$

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

$$= \frac{19.5}{5} - \frac{2.28 * 20}{5} - \frac{-1.67 * 24}{5}$$

$$\Rightarrow 2.796$$