Regression

* Regression analysis is said to be under supervised Learning
* it is a statistical method to model the relationship b/w dependent and independent variables
* The dependent variables are otherwise called as target
  Independent variables are otherwise called as predictors
* This type of regression models help us to understand how the value of the dependent variable is changing corresponding to an independent variable
* The regression model predicts continuous real values.
  eg: temperature, age, salary etc.

Types of Regression

1. linear regression
2. logistic regression
3. polynomial regression
4. support vector regression
5. decision tree

1. Linear regression

* it is a statistical method for predictive analysis
* if there lies only one i/p then such type of linear regression is called simple linear regression.
* if there exist more no. of i/p variables then such linear Regression is called Multiple linear regression.

* This shows Relationship b/t the independent variable which lie on the X-axis and the dependent variables which lie on the Y-axis

* Simple linear Regression formula

$$\boxed{y = B_0 + B_1 x + E}$$

where y = predicted value of dependent variables

B_0, B_1 = are coefficients

x = is an independent variable

E = Error occured (variation in our estimation)

$$\boxed{y = mx + c}$$

* y is said to the o/p value

x is said to the I/p value

m is slope of the line

c is given constant

Linear Regression

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{\Sigma (x - \bar{x})^2}$$

mean square error $\frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})^2$

Eg: consider the following data → $\boxed{\text{Question}}$

person	Rating Manual car X	Rating automation car Y
1	4	3
2	2	4
3	3	2
4	5	5
5	1	3
6	3	1

sol —

X	Y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x}).(y - \bar{y})$	$(y - \bar{y})^2$
4	3	1	0	1	0	0
2	4	-1	1	1	-1	1
3	2	0	-1	0	0	1
5	5	2	2	4	4	4
1	3	-2	0	4	0	0
3	1	0	-2	0	0	4

$\Sigma x = 18$ $\Sigma y = 18$ $\Sigma (x - \bar{x})^2 = 10$ $\Sigma (x - \bar{x})(y - \bar{y}) = 3$ $\Sigma (y - \bar{y})^2 = 10$

$$\bar{X} = \frac{\Sigma x}{T} = \frac{18}{6} = 3$$

$$\bar{y} = \frac{\Sigma y}{T} = \frac{13}{6} = 3$$

Q1. find the value of B_0 & B_1 with respect to ... model which best fits the given data

Sol

$$b_1 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2}$$

$$b_1 = \frac{3}{10} = 0.3$$

Consider $\hat{y} = \bar{y}$

$x = \bar{x}$

$\hat{y} = b_0 + b_1 x$

$3 = b_0 + 0.3(3)$

$b_0 = 3 - 0.9$

$$\boxed{b_0 = 2.1}$$

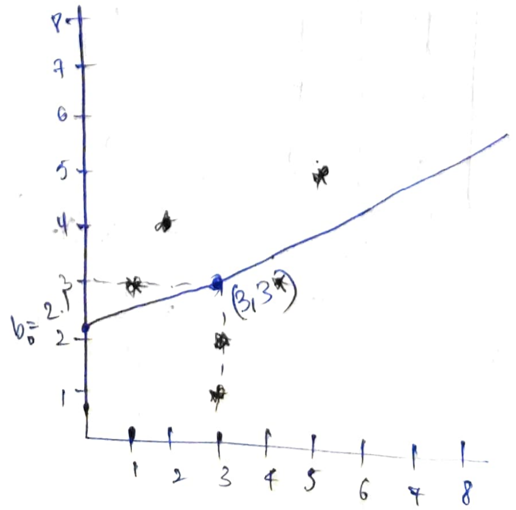
Q2. find the regression line, the best fit for given sample data.

Sol. $\hat{y} = b_0 + b_1 x$

$\hat{y} = 2.1 + 0.3x$

$b_0 = 2.1$

$(\bar{x}, \bar{y}) = (3, 3)$



Q3. Interpret & explain Equation of regression line

sol

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2$$

$$= \frac{10}{6}$$

Best fit 0 – 1.0

$$= 1.6$$

- The error generated for the given sample data exceeds 1

∴ The regression line is not a best fit for the given of data

- The error can be minimized by raising the no. of samples consider

Q4. if new person rates manual car as 4 then predict the rating of same person for automatic cars.

Sol –

$$\hat{y} = b_0 + b_1 x \qquad x = 4$$

$$= 2.1 + 0.3(4)$$

$$= 2.1 + 1.2$$

$$= 3.3$$