Regression

Linear Regression :-

Supervised learning is the learning in which a supervisor teaches or trains a machine using labeled data with correct answers. After that the machine is provided with new set of data so that supervised learning algo can produce the correct outcome from the labeled data with the help of training examples.

Regression analysis falls under supervised learning. Here system tries to predict a value for an input based on previous information.

Most Imp. charastics of Regression :

1) Responses obtained from the model are always quantitative.

2) The model can be constructed only if past data is available.

Simple linear regression is a statistical method that allows us to summarize and study the relationship b/w two continuous (quantitative) variables :

1) First variable, dentoted by $x \rightarrow$ predictor, explanatory or independent ~~value~~ variable

2) Second var, denoted by $y \rightarrow$ response, outcome, dependent var

Regression Model →

$$y_i = \beta_0 + \beta_1 x_i$$

$$\beta_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$\beta_0 = \frac{1}{n}\left(\Sigma y_i - \beta_1 \Sigma x_i\right) = \bar{y} - \beta_1 \bar{x}$$

Dataset of height + weight :-

Sr. No.	Height x (cm)	Weight y (kg)	A $(x_i - \bar{x})$	B $(y_i - \bar{y})$	A·B	A^2 $(x_i - \bar{x})^2$
1	151	63	-2.8	-2.3	6.44	7.84
2	174	81	20.2	15.7	317.14	408.04
3	138	56	-15.8	-9.3	146.94	249.64
4	186	91	32.2	25.7	827.54	1036.8
5	128	47	-25.8	-18.3	472.14	665.64
6	136	57	-17.8	-8.3	147.74	316.84
7	179	76	25.2	10.7	269.64	635.04
8	163	72	9.2	6.7	61.64	84.64
9	152	62	-1.8	-3.3	5.94	3.24
10	131	48	-22.8	-17.3	394.44	519.84
	$\bar{x} =$ 153.8	$\bar{y} =$ 65.3			$\Sigma =$ 2649.6	$\Sigma =$ 3927.6

$$\beta_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2649.6}{3927.6} = 0.6746$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 65.3 - (0.6746 \times 153.8) = -38.4551$$

β_1 and β_0 values are used to predict the weight of students.

For height = 170 cm

$$y = \beta_0 + \beta_1 x$$

$$y = -38.4551 + (0.6746 \times 170) = 76.23 \text{ kg}.$$

Steps to Establish a linear Regression:

A simple eg. of linear regression is to predict the weight of a person when his/her height is known.
Steps to create relationship b/w height & weight are as follows:

1) Carry out an experiment for gathering sample of observed values of height and corresponding weight.

2) Create a Relationship model.

3) Find the coefficients from the model created and establish the mathematical eq^n using these.

4) Get a summary of prediction to know the avg. error in prediction. This avg. error is also called residual.

5) Predict the weight of other people.

① Input data : Gather data in tabular form.

② create the Relationship model:

model can be stated as

$$y_i = \beta_0 + \beta_1 x_i \quad \text{--- (1)}$$

where y_i = known constant; it is value of response var. in i^{th} trial.

x_i = known constant; val. in i^{th} trial. → predictor

β_0, β_1 = Regression parameters.

→ To find "good" estimator of regression parameters β_0 & β_1, we use method of least squares.

In the observations (X_i, Y_i) for each case, method of least square consider the deviation of y_i from its expected value as given in eq^n (2)

$$y_i - (\beta_0 + \beta_1 x_i) \quad \leftarrow \text{(2)}$$

Method of least square ~~considers~~ requires that we consider the sum of the n squared deviations this is denoted by Q.

$$Q = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2 \quad \text{--- (3)}$$

According to the method of least squares, the estimators of β_0 and β_1 are those values b_0 and b_1 respectively, that minimize the criterion Q.

3) Derivation -

partially differentiating Q with respect to β_0 & β_1

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum x_i (y_i - \beta_0 - \beta_1 x_i) \quad — (4)$$

Equating the partial derivative of eq^n (4) to 0.

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \quad — (5)$$

let b_0 and b_1 are those values of β_0 and β_1 for which Q is minimum.

Simplifying & expanding Eq^n (5)

$$\sum y_i - nb_0 - b_1 \sum x_i = 0$$

$$\sum x_i y_i - b_0 \sum x_i - b_1 \sum x_i^2 = 0$$

Rearranging the terms and solving simultaneously for b_0 and b_1, we get.

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

$$b_0 = \frac{1}{n} \left(\sum y_i - b_1 \sum x_i \right) = \bar{y} - b_1 \bar{x}$$