

In Machine Learning, **supervised learning** is a type of learning where the model is trained on labelled data, meaning each input has a known output.

One common task in supervised learning is **regression**, which helps to understand the relationship between the independent and dependent variables, where the dependent variable (y) is a continuous numeric in nature such as price, salary or marks.

Linear Regression is a statistical technique that helps to understand the relationship between independent and dependent variables.

Linear Regression can help to understand how one unit change in the 'x' variable causes how many unit changes on the 'y' variable.

In Linear Regression, we have two types:

- 1) Simple Linear Regression
- 2) Multiple Linear Regression

A basic form of regression is **simple linear regression**, where the relationship between one independent variable and one dependent variable is modelled using a straight line, represented by the equation  $Y = \beta_0 + \beta_1 x$ , to predict the output value.

Where,

$\beta_0$ = y intercept (a, point where regression line has touch the y-axis)

$\beta_1$ = slope coefficient

The Assumptions for Simple Linear Regression are:

- 1) There should be no outliers.
- 2) X should show Linear relation with Y.
- 3) Y should follow approximate normal distribution.

Acceptance range for skewness (-0.5 to +0.5)/(-0.75 to +0.75)

Working for Simple Linear Regression:

- 1) Take the dataset with one input variable **X** and one output variable **Y**.
- 2) Plot **X vs Y** to check if a linear relationship exists.
- 3) Assume a straight-line equation:  $Y = \beta_0 + \beta_1 x$ .
- 4) Calculate the best values of **m (slope)** and **c (intercept)** using the **least squares method**.
- 5) Draw the regression line that best fits the data points.
- 6) Use the equation to **predict Y** for any new value of X.
- 7) Check model performance using error measures like **MSE** or **R<sup>2</sup> score**.