

Linear Regression

06 January 2024 12:53

Linear regression is used to predict the value of a variable based on other variables.

The variable which is to be predicted is called the **dependent variable** and the variable through which we are going to predict the value of other's variable is called **independent variable**.

Linear Regression works on datasets where the dependent variable is continuous.
Our aim is to find the best-fit line between the dependent and independent variable.

Best fit-line equation: $y = \beta x + \beta_0$
Where β = Slope of the line and β_0 = Intercept

Values of β and β_0 using OLS(Ordinary Least Squares Regression)

$$\beta = \frac{[\sum(x - \mu_x)(y - \mu_y)]}{\sum(x - \mu_x)^2}$$

where μ_x = mean of independent variable, μ_y = mean of dependent variable

$$\beta_0 = \mu_y - \beta \mu_x$$

Multiple Linear Regression is an extension of Linear Regression, where the dependent variables is dependent on several independent variables instead of single independent variable.

$$y = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad \text{where } x_1, x_2, x_3 \text{ are independent variables}$$

Univariate Regression deals with only one variable. It is only for data analysis and to find any patterns if exists.

Bivariate Regression deals with two variables like simple linear regression.

Multivariate Regression deals with more than two variables. Generally it has more than one outcomes.



Independent variable(x) on x-axis
Dependent variable(y) on y-axis