Linear Regression

Linear Regression is a supervised ML algorithm used for predictive analysis. Linear Regression shows a linear relationship between the output (y) ~~var~~ and input (x) variables.

Mathematically, linear regression is represented as -

$$y = a_0 + a_1 x + \varepsilon$$

$y \rightarrow$ output variable

$a_0 \rightarrow$ x-intercept

$a_1 \rightarrow$ slope (coefficient) of x

$\varepsilon \rightarrow$ random error

Types of Linear Regression :-

1. Simple Linear Regression - 1 input variable

2. Multiple Linear Regression - multiple input variables

Cost function :-

The goal of Linear Regression algorithm is to find the best fit line. The ~~best fit line~~ cost function (J) helps to find the best fit line by minimizing the Mean Squared Error (MSE).

$$J = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$J \rightarrow$ cost function

$n \rightarrow$ no. of observation

$\hat{y}_i \rightarrow$ predicted value

$y_i \rightarrow$ actual "

gradient descent is used to minimize the cost function (MSE)

☑ **R2 score :-**

R2 score (coefficient of determination) determines the goodness of fit. It is used to find the accuracy of model. The scale is $0 - 100$.

$$R2 \text{ score} = \frac{\text{explained variation}}{\text{total variation}}$$

☑ **Assumptions of Linear Regression :-**

1. Linear relationship between the features and target.
2. small or no multi-collinearity between the features.