

#Linear Regression

In the most simple words, Linear Regression is the supervised Machine Learning model in which the model finds the best fit linear line between the independent and dependent variable i.e it finds the linear relationship between the dependent and independent variable.

Linear Regression is of two types: Simple and Multiple. Simple Linear Regression is where only one independent variable is present and the model has to find the linear relationship of it with the dependent variable

Whereas, In Multiple Linear Regression there are more than one independent variables for the model to find the relationship.

Equation of Simple Linear Regression, where b_0 is the intercept, b_1 is coefficient or slope, x is the independent variable and y is the dependent variable.

linear regression 1

Equation of Multiple Linear Regression, where b_0 is the intercept, $b_1, b_2, b_3, b_4, \dots, b_n$ are coefficients or slopes of the independent variables $x_1, x_2, x_3, x_4, \dots, x_n$ and y is the dependent variable.

linear regression 2

A **Linear Regression** model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the **actual value** and **Predicted** value and the goal is to reduce this difference.

#Evaluation Metrics for Regression Analysis To understand the performance of the Regression model performing model evaluation is necessary. Some of the Evaluation metrics used for Regression analysis are:

1. **R squared or Coefficient of Determination:** The most commonly used metric for model evaluation in regression analysis is R squared. It can be defined as a Ratio of variation to the Total Variation. The value of R squared lies between 0 to 1, the value closer to 1 the better the model.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

where SSRES is the Residual Sum of squares and SSTOT is the Total Sum of squares

1. **Adjusted R squared:** It is the improvement to R squared. The problem/drawback with R2 is that as the features increase, the value of R2 also increases which gives the illusion of a good model. So the Adjusted R2 solves the drawback of R2. It only considers the features which are important for the model and shows the real improvement of the model. Adjusted R2 is always lower than R2.

$$R^2_{\text{adjusted}} = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

where

R^2 = sample R-square

p = Number of predictors

N = Total sample size.

1. **Mean Squared Error (MSE):** Another Common metric for evaluation is Mean squared error which is the mean of the squared difference of actual vs predicted values.

$$MSE = \frac{1}{n} \sum \underbrace{\left(y - \hat{y} \right)^2}_{\substack{\text{The square of the difference} \\ \text{between actual and} \\ \text{predicted}}}$$

1. **Root Mean Squared Error (RMSE):** It is the root of MSE i.e Root of the mean difference of Actual and Predicted values. RMSE penalizes the large errors whereas MSE doesn't.

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