

# Linear Regression

## Introduction

Regression is a relationship between input and output variables that is used for predicting the continuous variable.

**Linear Regression:** It is used to predict the value of a variable based on another variable. The variable you want to predict is called the **dependent variable**. The variable you are using to predict the other variable is called the **independent variable**.

In short, linear regression gives the relationship between the independent and dependent variable by making a line between them; however, that line must pass through most of the points.

## Linear Regression Equation

The equation of a line is given by:

$$y = mx + b$$

Where:

- $y$  is the dependent variable (the variable we want to predict).
- $x$  is the independent variable (the variable we use to predict  $y$ ).
- $m$  is the slope of the line (the change in  $y$  for a unit change in  $x$ ).
- $b$  is the y-intercept (the value of  $y$  when  $x$  is 0).

## Cost Function

The cost function is used to measure how well the linear regression model fits the data. The most common cost function for linear regression is the Mean Squared Error (MSE), which is defined as:

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

Where:

- $n$  is the number of data points.
- $y_i$  is the actual value of the dependent variable for the  $i$ -th data point.
- $\hat{y}_i$  is the predicted value of the dependent variable for the  $i$ -th data point.

The goal of linear regression is to minimize the cost function, which means finding the values of  $m$  and  $b$  that result in the smallest MSE.

## Value of $m$ and $b$

To find the values of  $m$  and  $b$  that minimize the cost function, we can use the following formulas:

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - m \cdot \bar{x}$$

Where:

- $\bar{y}$  is the mean of the dependent variable.
- $\bar{x}$  is the mean of the independent variable.