LINEAR REGRESSION

Linear Regression Theory

It is a statistical model that attempts to illustrate the relationship between independent and dependent variable, with a linear equation, if it is continuous.
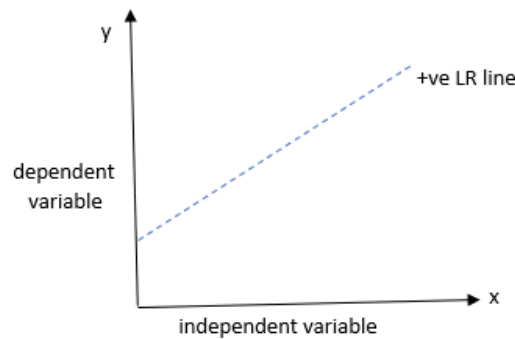
Line of linear regression is represented by the equation,
$$y = mx + c$$

where, y is the dependent variable, x is independent variable, m is slope of the line and c is y-intercept of the line.
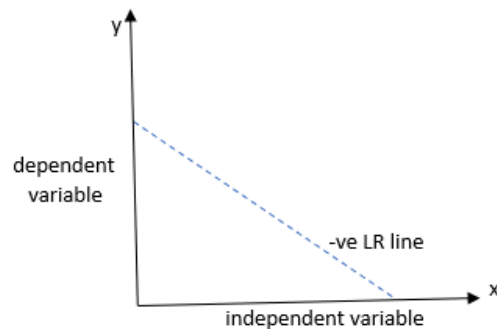
The slope can be calculated by,
$$m = \Sigma(x - \bar{x})(y - \bar{y}) / \Sigma(x - \bar{x})^2$$

where, $\bar{x}$ and $\bar{y}$ are the mean of x and y, respectively.

If the value of dependent variable increases with increase in the value of independent variable, we will get positive linear regression line.

If the value of dependent variable decreases with increase in the value of independent variable, we will get negative linear regression line.

Linear Regression with Python

Project Background
Data Source: https://www.kaggle.com/mirichoi0218/insurance

The dataset has been taken from Kaggle. The main objective of the analysis is to accurately forecast insurance charges. This dataset consists of 1338 rows and 7 columns.

Columns Description
· *age:* age of primary beneficiary
· *gender:* insurance contractor gender, female, male
· *bmi:* Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
· *children:* Number of children covered by health insurance / Number of dependents
· *smoker:* Smoking
· *region:* the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
· *charges:* Individual medical costs billed by health insurance