MrAkash920

Linear Regression

Regression: Regression analysis is a form of predictive modelling technique which investigates the relationship between dependent and independent variables.

Three major uses of regression analysis:

- Determine the strength of pattern
- Forecasting and effect
- Trend Forecasting

Linear Regression:

- It is the correlation between X and Y-axis.
- The data is modelled using straight line.
- Used for: Continuous variable.
- Output or Prediction is the values of the variable.
- Accuracy is measured by loss, R squared, Adjusted R Squared, etc.

Selection Criteria:

- Classification and Regression (Not good for classification all time)
- Data Quality should be good
- Computation Complexity (not much)
- Comprehensible and Transparent

Where Linear Regression use?

- Evaluating trends and sales estimates.
- Analyzing the impact of price changes.
- Assessment of risk in financial services and insurance domain.

Basic Structure:
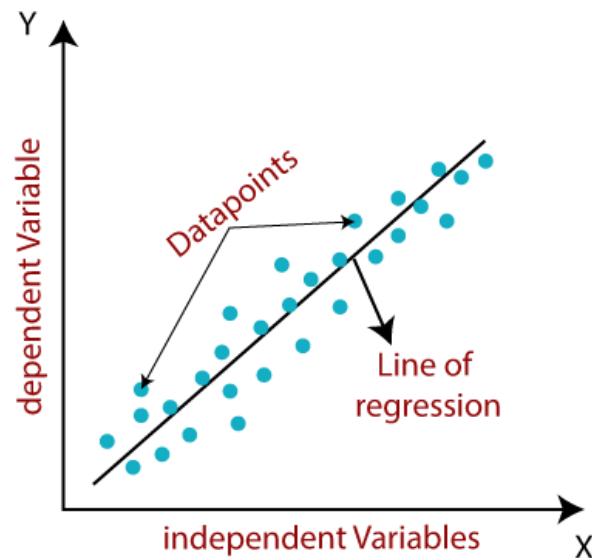
Mathematical Representation: $y = a_0 + a_1 x + \varepsilon$

Here,

Y= Dependent Variable (Target Variable)
X= Independent Variable (predictor Variable)
a_0= intercept of the line (Gives an additional degree of freedom)
a_1 = Linear regression coefficient (scale factor to each input value).
ε = random error

The values for x and y variables are training datasets for Linear Regression model representation.

Types of Linear Regression

Linear regression can be further divided into two types of the algorithm:

- o **Simple Linear Regression:**
  If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

- o **Multiple Linear regression:**
 If more than one independent variable is used to predict the value of a numerical

dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

Linear Regression Line

A linear line showing the relationship between the dependent and independent variables is called a **regression line**. A regression line can show two types of relationship:

- o **Positive Linear Relationship:**
  If the dependent variable increases on the Y-axis and independent variable increases on X-axis, then such a relationship is termed as a Positive linear relationship.

+ve line of regression
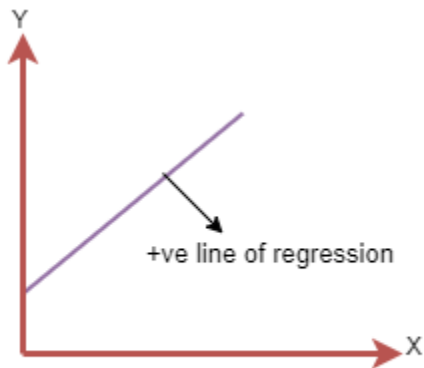
The line equation will be: $Y = a_0 + a_1x$

- o **Negative Linear Relationship:**
  If the dependent variable decreases on the Y-axis and independent variable increases on the X-axis, then such a relationship is called a negative linear relationship.

-ve line of regression
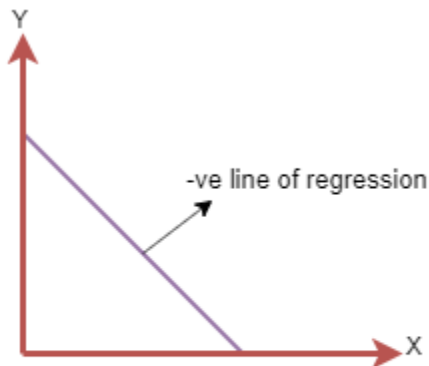
The line of equation will be: $Y = -a_0 + a_1x$

Model Performance:

The Goodness of fit determines how the line of regression fits the set of observations. The process of finding the best model out of various models is called optimization. It can be achieved by below method:

1. R-squared method:

- R-squared is a statistical method that determines the goodness of fit.
- It measures the strength of the relationship between the dependent and independent variables on a scale of 0-100%.
- The high value of R-square determines the less difference between the predicted values and actual values and hence represents a good model.
- It is also called a coefficient of determination, or coefficient of multiple determination for multiple regression.
- It can be calculated from the below formula:

$$\text{R-squared} = \frac{\text{Explained variation}}{\text{Total Variation}}$$