

LINEAR REGRESSION

By:
M. Ashish Reddy
DSWP -> Batch-5

What is a Regression

- In Regression, we plot a graph between the variables which best fit the given data points.
- The machine learning model can deliver predictions regarding the data. In naïve words, ***“Regression shows a line or curve that passes through all the data points on a target-predictor graph in such a way that the vertical distance between the data points and the regression line is minimum.”***
- It is used principally for prediction, forecasting, time series modeling, and determining the causal-effect relationship between variables.

Types of Regression models

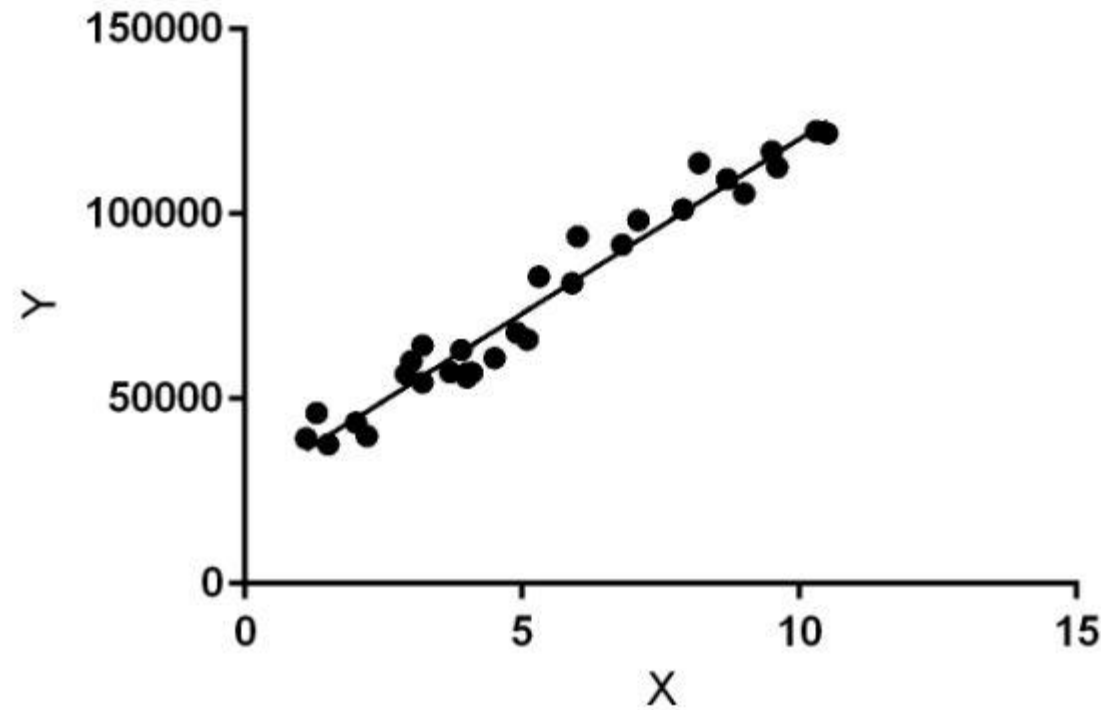
Linear
Regression

Polynomial
Regression

Logistics
Regression

Linear Regression

- **Linear Regression** is a machine learning algorithm based on **supervised learning**.
- It performs a **regression task**.
- Regression models a target prediction value based on independent variables.
- It is mostly used for finding out the relationship between variables and forecasting.
- Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.



- Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x).
- So, this regression technique finds out a linear relationship between x (input) and y(output).
- Hence, the name is Linear Regression.
- In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

$$y = mx + b \implies y = a_0 + a_1x$$

o calculate best-fit line linear regression uses a traditional slope-intercept form.

y= Dependent Variable.

x= Independent Variable.

a0= intercept of the line.

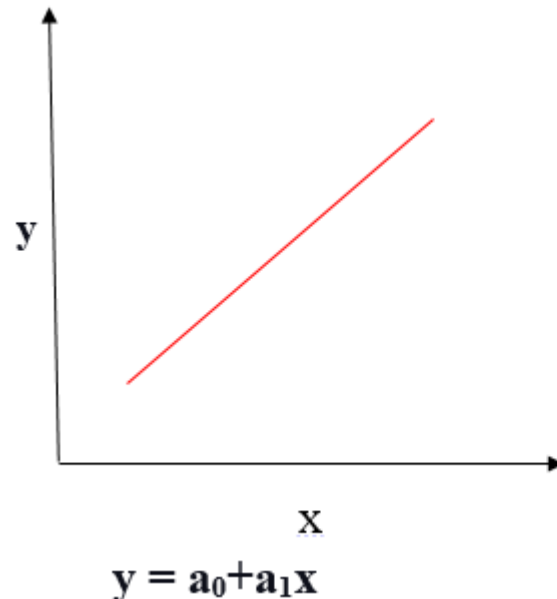
a1 = Linear regression coefficient.

Need of a Linear regression

- As mentioned above, Linear regression estimates the relationship between a dependent variable and an independent variable. Let's understand this with an easy example:
- Let's say we want to estimate the salary of an employee based on year of experience. You have the recent company data, which indicates that the relationship between experience and salary. Here year of experience is an independent variable, and the salary of an employee is a dependent variable, as the salary of an employee is dependent on the experience of an employee. Using this insight, we can predict the future salary of the employee based on current & past information.

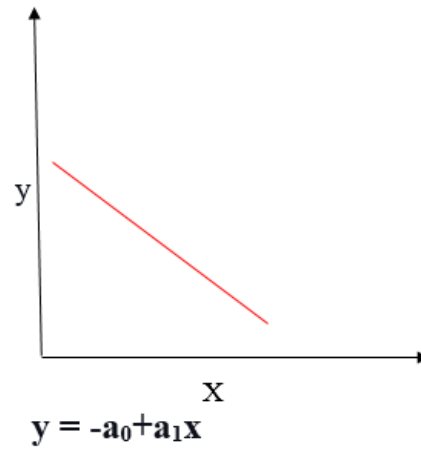
A regression line can be a Positive Linear Relationship or a Negative Linear Relationship.

- If the dependent variable expands on the Y-axis and the independent variable progress on X-axis, then such a relationship is termed a **Positive linear relationship**



Negative Linear Relationship

- If the dependent variable decreases on the Y-axis and the independent variable increases on the X-axis, such a relationship is called a negative linear relationship.



- The goal of the linear regression algorithm is to get the best values for a_0 and a_1 to find the best fit line. The best fit line should have the least error means the error between predicted values and actual values should be minimized.

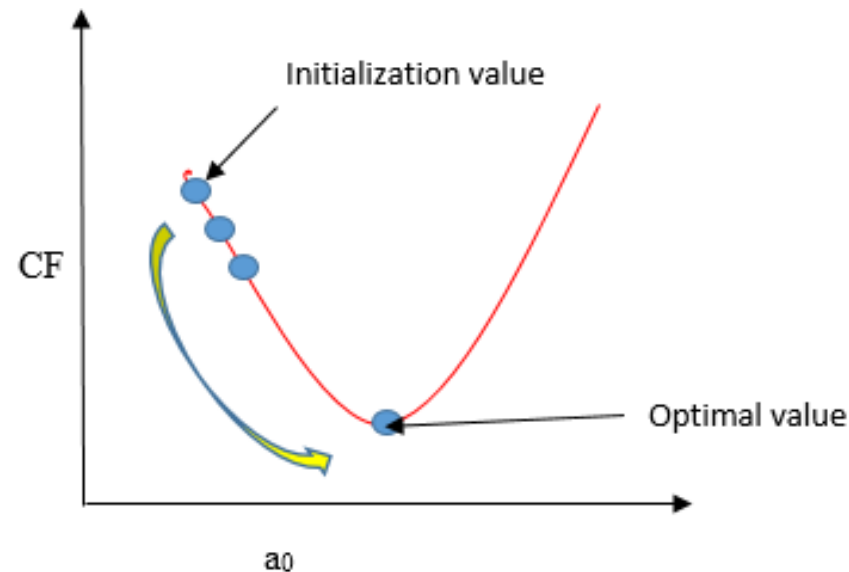
Cost function

- The cost function helps to figure out the best possible values for a_0 and a_1 , which provides the best fit line for the data points.
- Cost function optimizes the regression coefficients or weights and measures how a linear regression model is performing. The cost function is used to find the accuracy of the **mapping function** that maps the input variable to the output variable. This mapping function is also known as **the Hypothesis function**.
- In Linear Regression, **Mean Squared Error (MSE)** cost function is used, which is the average of squared error that occurred between the predicted values and actual values.
- *By simple linear equation $y=mx+b$ we can calculate MSE as:*
- *Let's y = actual values, y_i = predicted values*

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - (mx_i + b))^2$$

Gradient descent

- Gradient descent is a method of updating a_0 and a_1 to minimize the cost function (MSE). A regression model uses gradient descent to update the coefficients of the line ($a_0, a_1 \Rightarrow x_i, b$) by reducing the cost function by a random selection of coefficient values and then iteratively update the values to reach the minimum cost function.



Summary

- In Regression, we plot a graph between the variables which best fit the given data points. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis).
- To calculate best-fit line linear regression uses a traditional slope-intercept form. A regression line can be a Positive Linear Relationship or a Negative Linear Relationship.
- The goal of the linear regression algorithm is to get the best values for a_0 and a_1 to find the best fit line and the best fit line should have the least error.
- In Linear Regression, **Mean Squared Error (MSE)** cost function is used, which helps to figure out the best possible values for a_0 and a_1 , which provides the best fit line for the data points. Using the MSE function, we will change the values of a_0 and a_1 such that the MSE value settles at the minima.
- Gradient descent is a method of updating a_0 and a_1 to minimize the cost function (MSE)