**1.2) Linear Regression**:

It is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on continuous variable(s). Here, we establish relationship between independent and dependent variables by fitting a best line. This best fit line is known as regression line and represented by a linear equation Y= a \*X + b.

Before knowing what linear regression is, let us get ourselves accustomed to regression. Regression is a method of modeling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

Chart, scatter chart

Description automatically generated

Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. The red line in the above graph is referred to as the best fit straight line. Based on the given data points, we try to plot a line that models the points the best. The line can be modelled based on the linear equation shown below. y = a\_0 + a\_1 \* x • Y – Dependent Variable

* a – Slope
* X – Independent variable

The motive of the linear regression algorithm is to find the best values for a\_0 and a\_1.

Before moving on to the algorithm, let’s have a look at two important concepts you must know to better understand linear regression.

**Cost Function:**

The cost function helps us to figure out the best possible values for a\_0 and a\_1 which would provide the best fit line for the data points. Since we want the best values for a\_0 and a\_1, we convert this search problem into a minimization problem where we would like to minimize the error between the predicted value and the actual value.

A picture containing text, clock, watch

Description automatically generated

A picture containing text, clock

Description automatically generated

Cost function(J) of Linear Regression is the **Root Mean Squared Error (RMSE)** between predicted y value (pred) and true y value (y).

Minimization and Cost Function

We choose the above function to minimize. The difference between the predicted values and ground truth measures the error difference. We square the error difference and sum over all data points and divide that value by the total number of data points. This provides the average squared error over all the data points. Therefore, this cost function is also known as the Mean Squared Error (MSE) functions. Now, using this MSE function we are going to change the values of a\_0 and a\_1 such that the MSE value settles at the minima.