Linear Regression

Regression in simple terms is a method of modelling a target variable based on independent variables. The linear regression algorithm in machine learning is a supervised learning technique used to forecast and find out the relationship between variables.

General formula used to calculate the linear regression is

$$Y = a + bx$$

-Y is dependent variable whose value is dependent on the variable x.

-x is an independent variable.

-a is y-intercept

-b is coefficient of independent variable of x

Linear regression is a regression analysis in which independent variable has a linear relationship with dependent variable. The motive of linear regression is to find the best values for a and b.
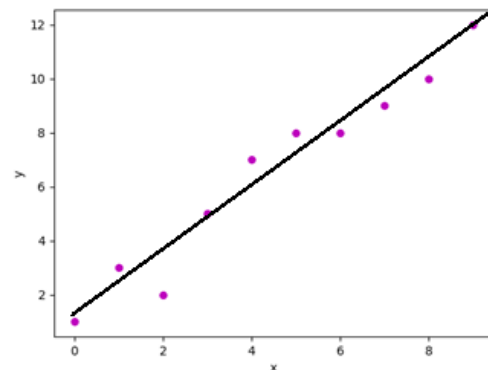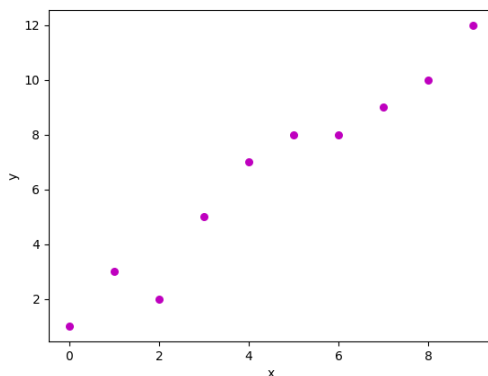
Important concepts in linear regression are:

Cost Function

The cost function is used to determine the best possible values for a and b to make

the best fit straight line for data with the minimal error.

For Linear Regression, **Mean Squared Error (MSE)** cost function is used, which is the average of squared error occurred between the predicted values and actual values.

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Where,

n=Total number of observation
yi = Actual value
predi =Predicted value.

Gradient descent

To update θ_1 and θ_2 values in order to reduce Cost function (minimizing MSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random a and b values and then iteratively updating the values, reaching minimum cost.