**Aim:** To obtain the best fit line over single feature scattered datapoints using Linear Regression

**IDE:** Google Colab

# Theory:

Linear regression is a method for determining the best linear relationship between two variables *X* and *Y*. If variables *X* and *Y* are uncorrelated, it is pointless embarking upon linear regression. However, if a reasonable degree of correlation exists between *X* and *Y* then linear regression may be a useful means to describe the relationship between the two variables. The usual approach is to use the *least-squares* method, which minimizes the squared difference between the actual data points and a straight line. Let [*xi,yi*], *i* = 1,2,3,….,*N* be the *N* pairs of data values of the variables *X* and *Y*. The straight-line relating *X* and *Y* is *y = mx + c*, where *m* and *c* are the gradient and constant values (to be determined) defining the straight line. Thus, *y(xi) - yi* is the difference between the line and data point *i* (see Fig. 1). Taking all the data points, we seek values of *m* and *c* that minimize the squared difference *SD*.



This is achieved by calculating the partial derivatives of *SD* with respect to *m* and *c* and finding the pair [*m*,*c*] such that *SD* is at a minimum.

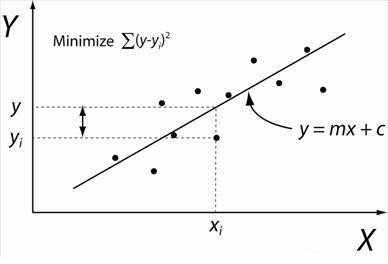


Figure 1: Illustration of Linear Regression. Linear least squares regression, the idea is to find the line y = mx + c that minimizes the mean squared difference between the line and the data points

# Batch Gradient Descent:

Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is basically used for updating the parameters of the learning model. Batch gradient descent which processes all the training examples for each iteration of gradient descent. But if the number of training examples is large, then batch gradient descent is computationally very expensive.

Let m be the number of training examples. Let n be the number of features.

**Algorithm for batch gradient descent :**

Let hθ(x) be the hypothesis for linear regression. Then, the cost function is given by:

Let Σ represents the sum of all training examples from i=1 to m.

Jtrain(θ) = (1/2m) Σ( hθ(x(i)) - y(i))2

Repeat {

θj = θj – (learning rate/m) \* Σ( hθ(x(i)) - y(i))xj(i)

For every j =0 …n

}

Where xj(i) Represents the jth feature of the ith training example. So if *m* is very large(e.g. 5 million training samples), then it takes hours or even days to converge to the global minimum.That’s why for large datasets, it is not recommended to use batch gradient descent as it slows down the learning.

# Pre Lab Exercise:

1. Explain the meaning of linear regression
2. Write three applications of linear regression
3. Write three advantages of linear regression
4. Write three limitations of linear regression

# Methodology:

1. Load the basic libraries and packages
2. Load the dataset
3. Analyse the dataset
4. Pre-process the data
5. Visualize the Data
6. Separate the feature and prediction value columns
7. Write the Hypothesis Function
8. Write the Cost Function
9. Write the Gradient Descent optimization algorithm
10. Apply the training over the dataset to minimize the loss
11. Find the best fit line to the given dataset
12. Observe the cost function vs iterations learning curve

# Program (Code):

To be attached with

# Results:

To be attached with

1. Datapoints scattering (without best fit line)
2. Scatter plot showing the Best fit line in the last iteration of training

# Observation and Result Analysis:

1. Nature of the dataset
2. During Training Process
3. After the training Process

# Post Lab Exercise:

1. What are the major assumptions considered in linear regression
2. Why MSE is used instead of MAE for calculating the loss function
3. How can the behaviour of outliers be understood while dealing with the unseen dataset
4. Derive the Normal Equation for the Linear Regression.