

Sumon Singh  
30.4.22

AIM :- Write a program to implement Linear regression algorithm

### 1) Dataset Description:

A) Name of dataset: - fish.csv

B) Description: This dataset consist of data of various kind of fishes. The data are weight, Length, Height etc. This dataset is a numerical dataset. The data set is available in Kaggle website.

C) Size: The dataset has 159 rows and 7 columns.

D) Attributes: This dataset has 6 features.

a) Species: Name of the fish

b) ~~Length~~ Length1: Vertical Length of fish

c) Length 2: diagonal Length of fish

d) Length 3: Cross Length of fish

e) Height: height of fish

f) width: width of fish

E) Label: The target of this dataset is weight of the fish which is a numeric data in Gram.



Sumon Singh  
30.4.22

## 2) Model Evaluation :-

A) Name :- Multiple linear regression

B) Type :- It's a supervised learning model which is used for regression type problem.

C) Algorithm :-

Input :- Data samples of training dataset

Step 1: Create n-dimension graph if the input dataset has n columns

Step 2: Plot all the attributes data points on the graph and generate an equation of straight line such that the output and the target data has the minimum error that means generate a straight line of best fit.

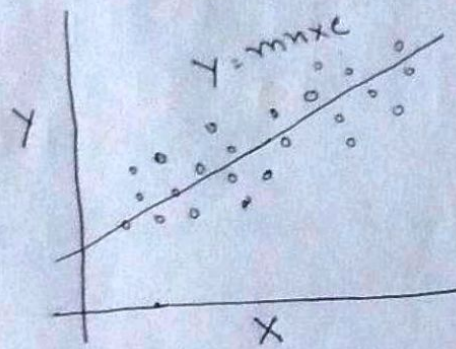
So, the equation will be :-

$$m_1x_1 + m_2x_2 + \dots + b = y$$

Step 3: Based on the generated best fit line and the values of intercepts and slope, new datapoints are tested.



D) Draw :-



X → feature  
Y → target

3) Result Analysis :-

a)  $r^2$  score :- 0.89, that means 89% of <sup>predicted</sup> target data is almost explained by the features correctly.

b) mean absolute error :- 94.667, that means the absolute error average of predicted and target data is 94.667

c) mean squared error :- 12283.9287, that means average square error of predicted and target data is 12283.9287