EXPERIMENT – 4

Implementation of Linear Regression

CLASS: TE CMPN A **PID:182027**
NAME: Rebecca Dias **ROLL NO.: 19**

Aim: Implement Linear Simple Regression model

Theory:

- **Explain Regression**
  Regression analysis is a statistical method that helps us to analyse and understand the relationship between two or more variables of interest. The process that is adapted to perform regression analysis helps to understand which factors are important, which factors can be ignored and how they are influencing each other.

- **Explain types of linear Regression**
  Linear Regression is generally classified into two types:
  1. Simple Linear Regression

     In Simple Linear Regression, we try to find the relationship between a single independent variable (input) and a corresponding dependent variable (output). This can be expressed in the form of a straight line.

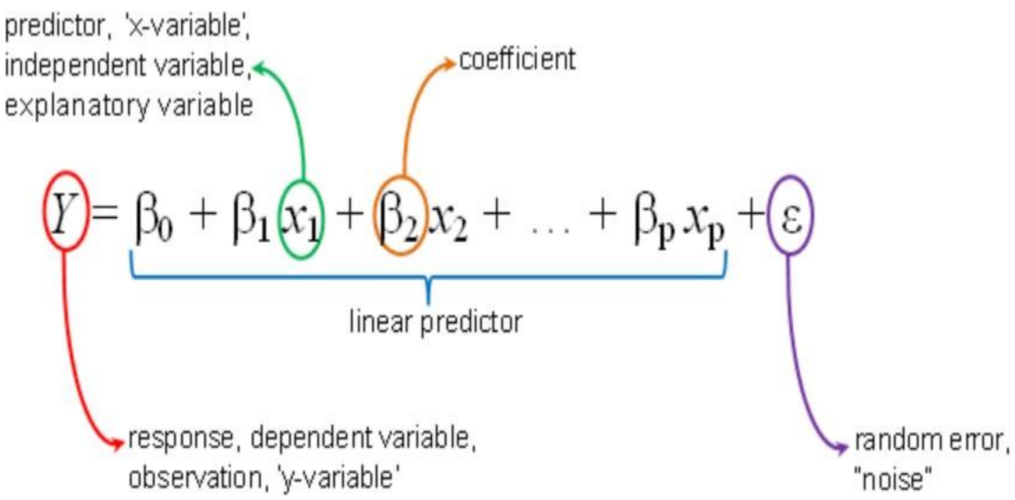
$$Y = \beta_0 + \beta_1 X + \epsilon$$

     - Y represents the output or dependent variable
     - β0 and β1 are two unknown constants that represent the intercept and coefficient (slope) respectively.
     - ε (Epsilon) is the error term.

2. Multiple Linear Regression

In Multiple Linear Regression, we try to find the relationship between 2 or more independent variables (inputs) and the corresponding dependent variable (output).

The independent variables can be continuous or categorical. The equation that describes how the predicted values of y is related to p independent variables is called as Multiple Linear Regression equation:



- **Differentiate in Regression and classification**

Regression Algorithm	**Classification Algorithm**
In Regression, the output variable must be of continuous nature or real value.	In Classification, the output variable must be a discrete value.
The task of the regression algorithm is to map the input value (x) with the continuous output variable(y).	The task of the classification algorithm is to map the input value(x) with the discrete output variable(y).
Regression Algorithms are used with continuous data.	Classification Algorithms are used with discrete data.
In Regression, we try to find the best fit line, which can predict the output more accurately.	In Classification, we try to find the decision boundary, which can divide the dataset into different classes.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc.
The regression Algorithm can be further divided into Linear and Non-linear Regression.	The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier.

- **Manually solve the problem given**

x	y	xy	x^2	y^2
3	30	90	9	900
8	57	456	64	3249
9	64	576	81	4096
13	72	936	169	5184
3	36	108	9	1296
6	43	258	36	1849
11	59	649	121	3481
21	90	1890	141	8100
1	20	20	1	400
16	83	1328	256	6889
Σ 91	554	6311	1187	35444

$$a = \frac{\Sigma y \Sigma x^2 - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$= \frac{83297}{3589}$$

$$= 23.20897$$

$$b = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{12696}{3589}$$

$$= 3.537476$$

For prediction

$$y = a + bx$$

let $x = 15$

$$\therefore y = a + 15b$$

$$= 23.20897 + 15 \times 3.537476$$

$$= 76.2716$$

Implementation:

- **Develop a model of linear regression for the above data set given (take values of data set at run time)**

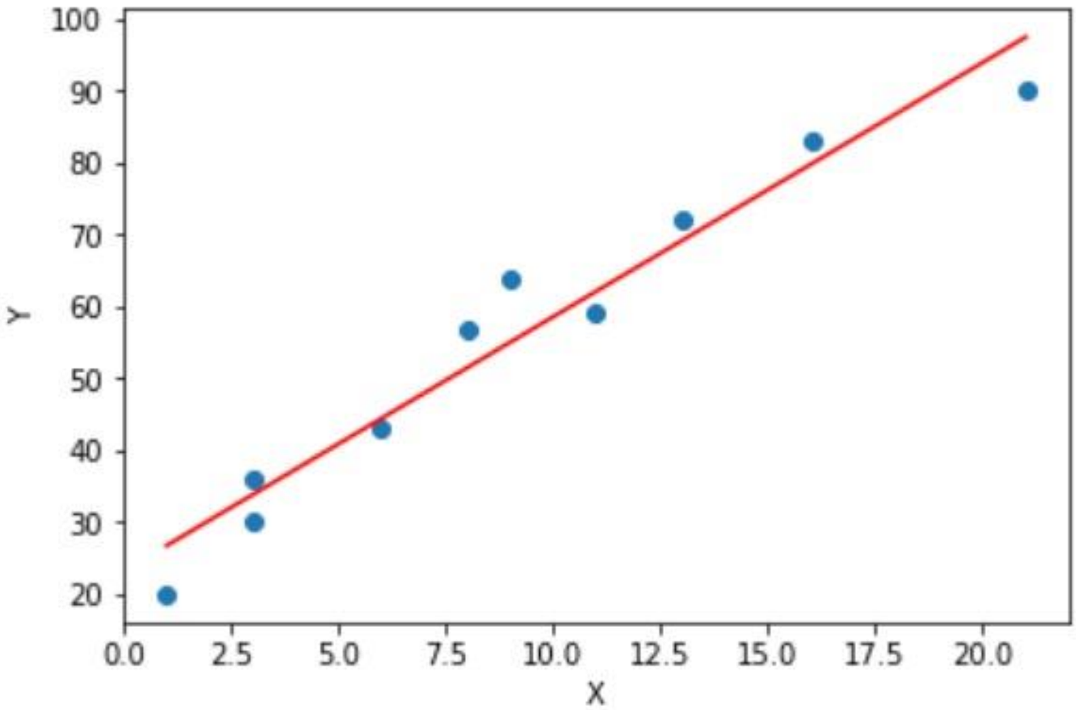
```
import numpy as np
import matplotlib.pyplot as plt
def predict(b0,b1,x):
    return b1+b0*x
def coefficient(x,y):
    xm=np.mean(x)
    ym=np.mean(y)
    X2=[(x[i])* (x[i]) for i in range(len(x))]
    Y2=[(y[i])* (y[i]) for i in range(len(y))]
    XY=[(y[i])* (x[i]) for i in range(len(y))]
    num1=sum(y)*sum(X2)-sum(x)*sum(XY)
    den=len(x)sum(X2)-sum(x)*2
    num2=len(x)*sum(XY)-sum(x)*sum(y)
    b1=num1/den
    b0=num2/den
    return b0,b1

x=list(map(int,input("Enter x (Year of Experience) : ").split()))
y=list(map(int,input("Enter y (Salary in $100): ").split()))

b0,b1=coefficient(x,y)
print("b0 : ",b0,"\n","b1 : ",b1)
n=int(input("Enter value : "))
print("Expected salary : ",predict(b0,b1,n))
x1=np.linspace(min(x),max(x),10)
y1=b1+b0*x1
plt.scatter(x,y)
plt.xlabel("X")
plt.ylabel("Y")
plt.plot(x1,y1,'-r')
plt.show()
```

- **Plot the linear regression line.**
- **For new tuple of x try to predict the value of y using the model. (take any new value of years of experience at run time and predict salary)**

```
Enter x (Year of Experience) : 3 8 9 13 3 6 11 21 1 16
Enter y (Salary in $100): 30 57 64 72 36 43 59 90 20 83
b0 :  3.5374756199498467
 b1 :  23.208971858456394
Enter value : 15
Expected salary :  76.2711061577041
```



classmate

Date
Page

Conclusion:

Summary of Experiment

- In this experiment, we learnt about the use and implementation of the different types of linear Regression
- We learnt about the difference between regression and classification
- We also solved an example to verify the model created and plotted the linear regression plot

Importance of Experiment

- Linear regression models are relatively simple and provide an easy to interpret mathematical formula that can generate predictions. Linear regression can be applied to various areas in business and academic study
- Linear regression is a long-established statistical procedure, the properties of linear-regression models are well understood and can be trained very fast

Applications of Experiment

- Used in Business to evaluate trends and make estimates or forecasts
- Used to analyse the marketing effectiveness, pricing and promotions on sales of a product
- Can be used to assess risk in financial services or insurance domain.