# Machine Learning

## Experiment 1

Ayush Jain 60004200132 B3

**Aim :** To implement univariate linear regression in python with and without using any inbuilt function.

**Theory:**
Linear regression is a statistical method that is used to analyze the relationship between two continuous variables. In simple linear regression, the goal is to find a linear relationship between the predictor variable (X) and the response variable (Y), which can be expressed by the equation:

$Y = a + bX$

where Y is the predicted value of the response variable, X is the value of the predictor variable, a is the intercept or the value of Y when X is 0, and b is the slope or the change in Y for every unit increase in X.

The process of linear regression involves finding the best values for the intercept and slope that minimize the sum of the squared differences between the actual values of Y and the predicted values of Y based on the values of X. This is done using a technique called least squares regression.

Linear regression can be used to make predictions or to model the relationship between variables in order to gain insights and make decisions. It is commonly used in fields such as finance, economics, and social sciences.

Let's take an example of a simple linear regression, where we want to predict the salary of an employee based on their years of experience. We have a dataset of 10 employees with their corresponding years of experience and salaries:

| **Years of Experience** | **Salary (in thousands)** |
|---|---|
| **1** | **20** |
| **2** | **22** |

| Years of Experience | Salary (in thousands) |
|---|---|
| 3 | 25 |
| 4 | 28 |
| 5 | 30 |
| 6 | 32 |
| 7 | 35 |
| 8 | 38 |
| 9 | 40 |
| 10 | 42 |

We can plot this data on a scatter plot to see if there is a linear relationship between years of experience and salary:

scatter plot of years of experience and salary

From the scatter plot, we can see that there appears to be a positive linear relationship between years of experience and salary, meaning that as years of experience increase, salary tends to increase as well.

We can then perform linear regression on this data to find the line of best fit that represents the relationship between years of experience and salary. The equation for the line of best fit is:

Salary = 18.98 + 4.16*Years of Experience

The line of best fit represents the predicted salary for a given number of years of experience. For example, if an employee has 7 years of experience, we can predict their salary to be approximately $35,000.

Linear regression can be a powerful tool for analyzing and predicting relationships between variables. However, it's important to keep in mind that correlation does not always imply causation, and there may be other factors that influence the relationship between variables.

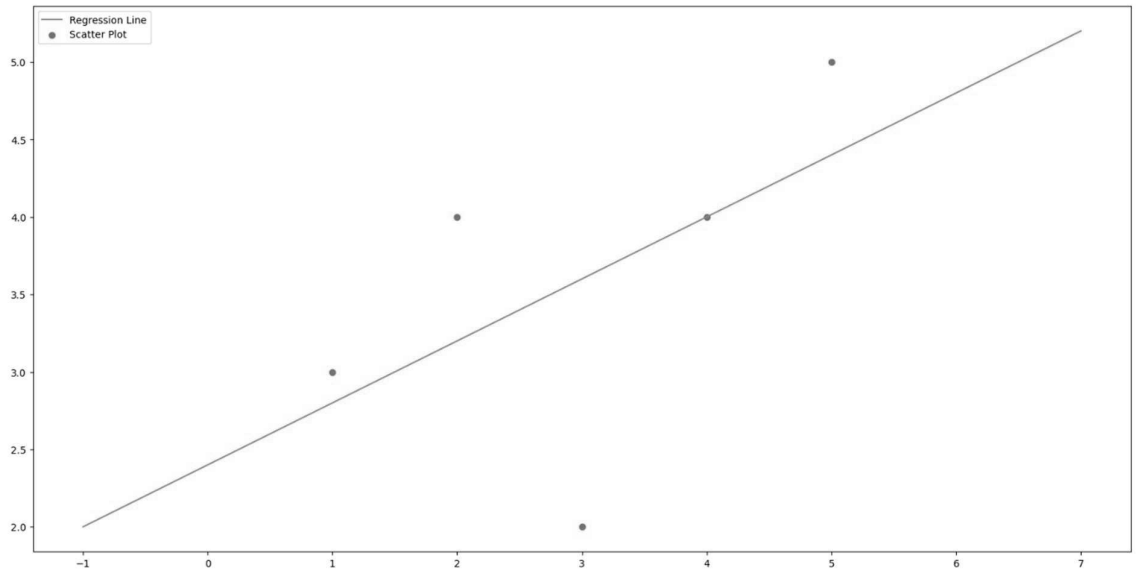
**Code without inbuilt function:**

```
x = [1,2,3,4,5]
y = [3,4,2,4,5]

import numpy as np
```

```
import pandas as pd
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = (20.0, 10.0)

X = np.array([1,2,3,4,5])#data['Head Size(cm^3)'].values
Y = np.array([3,4,2,4,5])#data['Brain Weight(grams)'].values

mean_x = np.mean(X)
mean_y = np.mean(Y)

n = len(X)

numer = 0
denom = 0
for i in range(n):
    numer += (X[i] - mean_x) * (Y[i] - mean_y)
    denom += (X[i] - mean_x) ** 2
    w0= numer / denom
    w1 = mean_y - (w0 * mean_x)

print(w0, w1)

max_x = np.max(X) + 2
min_x = np.min(X) - 2

# Calculating line values x and y
x = np.linspace(min_x, max_x, 1000)
y = w1 + w0 * x

plt.plot(x, y, color='#52b920', label='Regression Line')
plt.scatter(X, Y, c='#ef4423', label='Scatter Plot')


plt.legend()
plt.show()
```

**Output:**

Value of w0, w1: 0.4 2.4

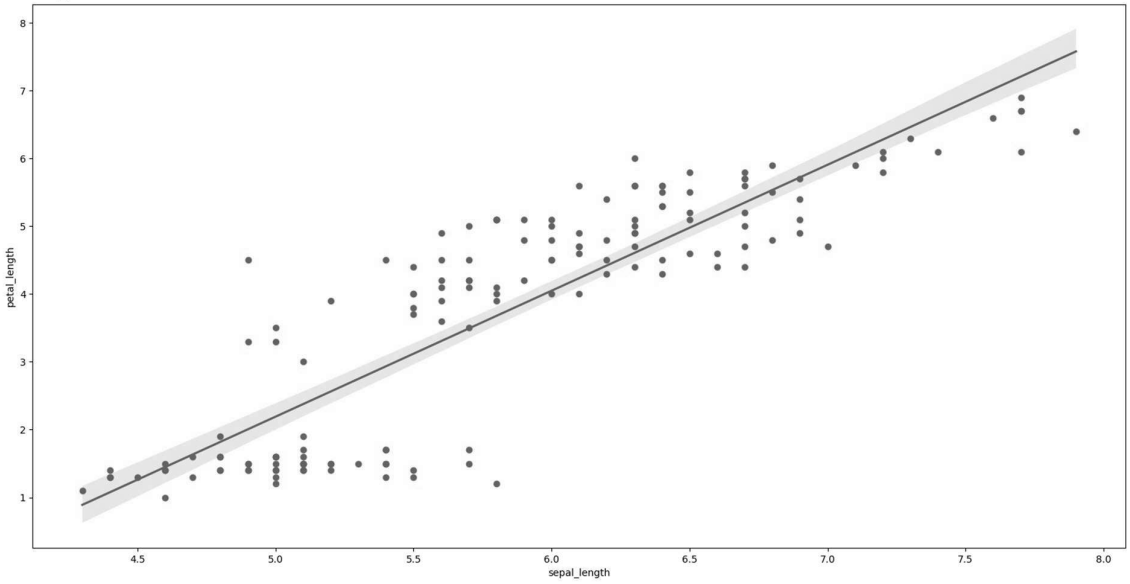



**Code with inbuilt function:**

```
import pandas as pd
import numpy as np
from sklearn.datasets import load_iris
from sklearn.linear_model import LinearRegression

iris = load_iris()

df = pd.DataFrame(data= np.c_[iris['data'], iris['target']],
                  columns= iris['feature_names'] + ['target'])

X = df[['sepal length (cm)']]
y = df['petal length (cm)']

reg = LinearRegression()

reg.fit(X, y)

print('Slope:', reg.coef_[0])
print('Intercept:', reg.intercept_)

new_X = [[6.5]]
print('Predicted Y:', reg.predict(new_X))

import matplotlib.pyplot as plt
import seaborn as sns

# load the Iris dataset
```

```
iris = sns.load_dataset('iris')

sns.scatterplot(x='sepal_length', y='petal_length', data=iris)
sns.regplot(x='sepal_length', y='petal_length', data=iris)
plt.show()
```

**Output:**




**Conclusion:** We successfully implemented Linear Regression.