Experiment 4

Shashwat Shah
60004220126
C2-2 Div B

Aim: Implementation of linear Regression for single variate and Multivariate

Theory: Linear regression is a fundamental statistical method used for modeling the relationship between one or more independent variables and a dependent variable. It assumes a linear relationship between the variables aiming to find the best-fitting line that maximizes the sum of squared differences between observed and predicted values.

It used in finace to predict stock prices. It is used to extract insights from data and make informed predictions in various domains.

Single variate linear regression

we model the relationship between a single independent and a dependent variable. The goal is to find the best-fit line represented by $y = mx + b$ when m is the slope & b is the y-interrupt.

Multivariate linear regression

we model a relationship between multiple variables. The Model equation becomes: $y = b_0 + b_1 x_1 b_2 x_2 + b_p x_p$

Where bo is the y intercept and b_1, b_2, b_p are the co-efficients. The goal is to minimize the sum of squared differences. This approach accomodates the complexity of real world scenarios with multiple predictions allowing for more nuanced understanding of the relationship between variables.

<u>Conclusion</u>: Hence, we implemented single variate & multivariate linear regression.

NAME:Shashwat Shah

SAPID:60004220126

DIVISION:C2

DMW EXP 4

Aim: Implementation of Linear Regression for Single Variate and Multi-variate.

Theory:

Simple Linear Regression (Single Variate): In simple linear regression, there is a single independent variable (predictor) and a single dependent variable (response). The goal is to establish a linear relationship between them. The fundamental equation of simple linear regression is:

Y=aX+b Where:

- Y represents the dependent variable.
- X represents the independent variable.
- a is the slope of the regression line, indicating the change in Y for a one-unit change in X.
- b is the y-intercept, representing the value of Y when X is zero.

The objective of simple linear regression is to find the values of a and b that minimize the sum of squared differences between the observed data points and the values predicted by the linear equation. This is typically achieved using the method of least squares.

Simple linear regression is particularly useful for exploring and modeling relationships between two variables when you suspect that a linear relationship exists.

Multiple Linear Regression (Multi-Variate): Multiple linear regression extends the concept of simple linear regression to situations where there are multiple independent variables influencing a single dependent variable. The equation for multiple linear regression is as follows:

Y=a1X1+a2X2+…+anXn +b Where:

- Y is the dependent variable.
- X1, X2..,Xn are the independent variables.

- a1,a2,…,an are the coefficients (slopes) associated with each independent variable.
- b is the y-intercept.

The goal in multiple linear regression is to estimate the coefficients (a1,a2,…,an) in a way that best fits the observed data. This is also done using the least squares method, but it involves solving a system of equations to determine the optimal coefficients.

Multiple linear regression is a powerful tool for modeling real-world phenomena where the outcome depends on multiple factors simultaneously. It is widely used in fields such as economics, social sciences, and engineering to make predictions, understand relationships, and assess the significance of different variables.

Part A: Univariate Regression

Fuel_Consumption_City Vs CO_2 Emissions:

Cell 1:

```
import numpy as np import
pandas as pd
 df =
pd.read_csv('/content/Fuel.csv')
```

Output: Imported Successfully

Cell 2:

```
from sklearn.model_selection import train_test_split
X = df[["Fuel_Consumption_City"]] y =
df[["CO2_Emissions"]]

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=0)
```

Cell 3:

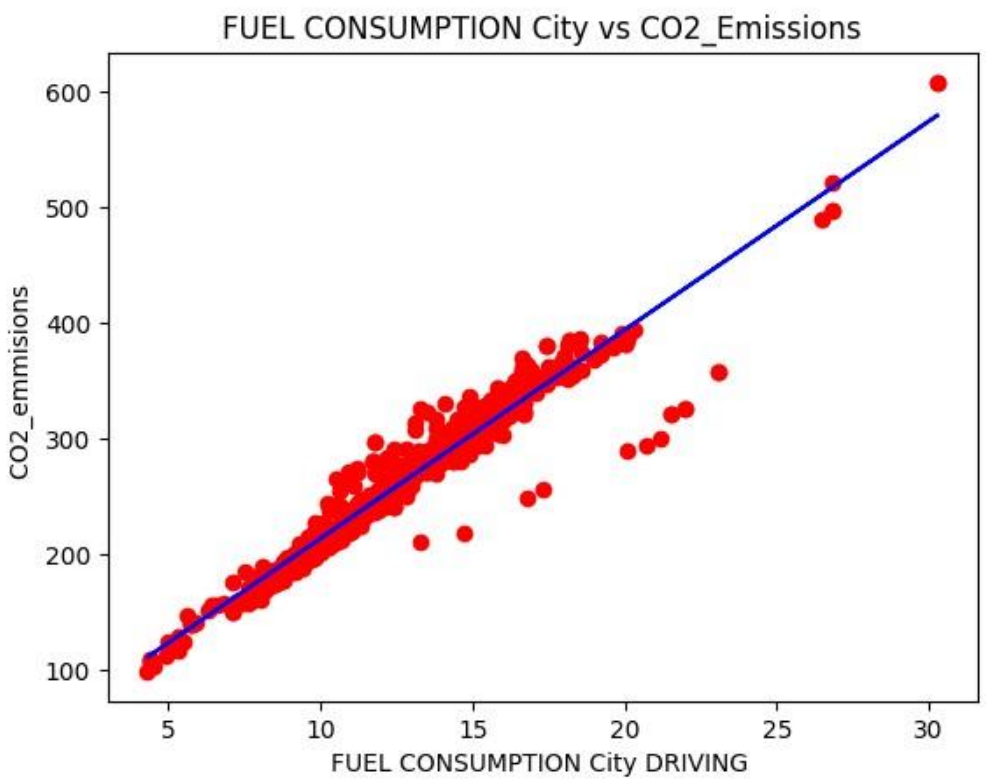
```
from sklearn.linear_model import LinearRegression
reg = LinearRegression().fit(X, y) y_pred =
reg.predict(X_test) accuracy = reg.score(X_test,
y_test) print(accuracy)
```

Output: `0.9455450318076177`

Cell 4:

```
import matplotlib.pyplot as plt
plt.scatter(X_train, y_train, c = 'r')
plt.plot(X_train, reg.predict(X_train), c = 'b')
plt.title('FUEL CONSUMPTION City vs CO2_Emissions')
plt.xlabel("FUEL CONSUMPTION City DRIVING ")
plt.ylabel("CO2_emmisions ") plt.show()
```

Output:



Part B: Multivariate Regression

Fuel_Consumption_City , Engine Size(L),Cylinders Vs CO_2 Emissions:

Cell 1:

```
x_ = df[['Engine Size(L)','Cylinders','Fuel_Consumption_City']] y_
= df[['CO2_Emissions']]
 x_train, x_test, Y_train, Y_test = train_test_split(x_,
y_, test_size=0.2, random_state=0) reg =
LinearRegression().fit(x_, y_) y_pred = reg.predict(x_test)
accuracy = reg.score(x_test, Y_test) print(accuracy)
```

Output: `0.9443478876839433`

Conclusion:

In this experiment, we implemented linear regression for both single variate and multi-variate datasets using Python sklearn library and its functions. We observed that:

- Single variate regression helps us understand and predict relationships between one independent variable and a dependent variable.
- Multi-variate regression extends this to multiple independent variables, offering a more comprehensive analysis.