**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: B.Tech CSDS(311)/B.Tech-CSBS

**Course: Machine Learning**

**Experiment No.02**

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical.)***

|  |  |
| --- | --- |
| Roll No.: C026 | Name: Anirbaan Ghatak |
| Class : B | Batch : B1 |
| Date of Experiment: 01/08/2023 | Date of Submission: 04/08/2023 |
| Grade : |  |

**B.1 Task1**

#Aim: Implementation of Simple Regression

#Name: Anirbaan Ghatak

#Roll no.: C026

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import r2\_score, mean\_squared\_error

df = pd.read\_csv('USA\_Housing.csv')

df.head(10)

df.info()

df.columns

sns.heatmap(df.corr(), annot=True)

X = df[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',

       'Avg. Area Number of Bedrooms', 'Area Population']]

Y = df['Price']

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.40, random\_state=101)

lm = LinearRegression()

lm.fit(X\_train, Y\_train)

coeff\_df = pd.DataFrame(lm.coef\_,X.columns,columns=['Coefficient'])

print(coeff\_df)

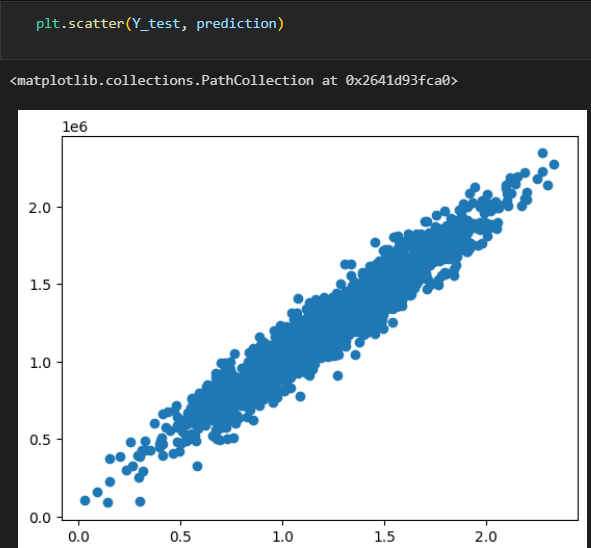
prediction = lm.predict(X\_test)

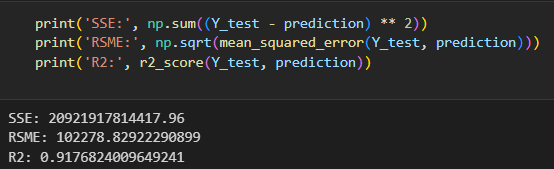
plt.scatter(Y\_test, prediction)

print('SSE:', np.sum((Y\_test - prediction) \*\* 2))

print('RSME:', np.sqrt(mean\_squared\_error(Y\_test, prediction)))

print('R2:', r2\_score(Y\_test, prediction))

****

****

**B.2 Conclusion:**

*Read about the types of regression Linear, logistic, polynomial. Understood the concept of simple linear regression and implemented the concepts learned using the given task in python.*

**Question of Curiosity:**

1. Consider the following set of points: {(-2, -1) , (1 , 1) , (3 , 2)}  
   a) Find the least square regression line for the given data points.

To find the least square regression line for the given data points:

y = mx + b

where,

m = (nΣ(xy) - ΣxΣy) / (nΣ(x^2) - (Σx)^2)

b = (Σy - mΣx) / n

For (-2, -1): x = -2, y = -1, x^2 = 4, xy = 2

For (1, 1): x = 1, y = 1, x^2 = 1, xy = 1

For (3, 2): x = 3, y = 2, x^2 = 9, xy = 6

Therefore,

Σx = -2 + 1 + 3 = 2

Σy = -1 + 1 + 2 = 2

Σxy = (-2 \* -1) + (1 \* 1) + (3 \* 2) = 1 + 1 + 6 = 9

Σ(x^2) = (-2)^2 + 1^2 + 3^2 = 4 + 1 + 9 = 14

n = 3

Thus,

m = (3 \* 9 - 2 \* 2) / (3 \* 14 - 2^2)

m = (24 - 4) / (42 - 4)

m = 20 / 38

m = 10 / 19 = 0.6052 (approx.)

and,

b = (2 - 0.5263 \* 2) / 3

b = (2 - 1.0526) / 3

b = 0.9474 (approx.)

So, the regression line that best fits the given data points is `y = 0. 6052x + 0.9474`.