UNIVERSITY OF MUMBAI

DEPARTMENT OF COMPUTER SCIENCE

M.Sc. Data Science Semester – II

Roll No :	17
Name of Student :	Shruti Soma Gawas
Subject :	Artificial Intelligence and Machine Learning Lab
Assignment no.:	2
Date:	31-03-2023

Q. 1) Model the salary prediction problem using linear regression algorithm**.**
Use the data set:
https://drive.google.com/file/d/1pZ-xIpt2qjvb5an9UpEGXjuF9lXAY2Ey/view?usp=sharing
<submit pdf of notebook>

◆◆◆◆◆

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [2]:

```
df=pd.read_csv('Salary_Data.csv')
```

In [3]:

```
df.head()
```

Out[3]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

In [4]:

```
x=df.iloc[:, :-1].values
y=df.iloc[:, 1].values
```

In [5]:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test=train_test_split(x,y,test_size=0.2,)
```

In [6]:

```
x_train.shape
```

Out[6]:

```
(24, 1)
```

In [7]:

```
y_train.shape
```

Out[7]:

```
(24,)
```

In [8]:

```
from sklearn.linear_model import LinearRegression
linreg=LinearRegression()
linreg.fit(x_train,y_train)
print ('Coeff:', linreg.coef_)
print ('Intercept:',linreg.intercept_)
```

```
Coeff: [9416.71369298]
Intercept: 26093.80805651524
```

In [9]:

```
y_pred= linreg.predict(x_test)
```

In [10]:

```
y_pred
```

Out[10]:

```
array([ 54343.94913546, 108019.21718545,  46810.57818107,  56227.29187405,
        72235.70515212,  63760.66282844])
```

In [11]:

```
y_test
```

Out[11]:

```
array([ 60150., 109431.,  39891.,  64445.,  67938.,  55794.])
```

In [12]:

```
plt.scatter(x_test,y_test, color='red')
plt.plot(x_train, linreg.predict(x_train),color='blue')
plt.title("Years of expereicne vs Salary")
plt.xlabel("Years of exp.")
plt.ylabel("Salary")
plt.show()
```

In [14]:

```
plt.scatter(x_train,y_train, color='red')
plt.plot(x_train, linreg.predict(x_train),color='blue')
plt.title("Years of expereicne vs Salary")
plt.xlabel("Years of exp.")
plt.ylabel("Salary")
plt.show()
```