Experiment No. - 3

Student Name: Banpreet Singh   **UID:** 20BCS2363

Branch: CSE   **Section/Group:** 20BCS_MM808/A

Semester: 5   **Date of Performance:** 25 Aug 2022

Subject Name: Machine Learning Lab   **Subject Code:** 20CSP-317

1. Aim/Overview of the practical:

Implement linear regression on any data set.

2. Software Used:

- Google Colab (https://colab.research.google.com/) or
- Jupyter Notebook

3. Code:

```python
# Banpreet Singh
# 20BCS2363
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

dataset = pd.read_csv('/content/Salary_Data.csv')
dataset.head()

# data preprocessing
X = dataset.iloc[:, :-1].values  #independent variable array
y = dataset.iloc[:,1].values  #dependent variable vector

# splitting the dataset
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=1/3,random_state=0)

# fitting the regression model
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train) #actually produces the linear eqn for the data
```
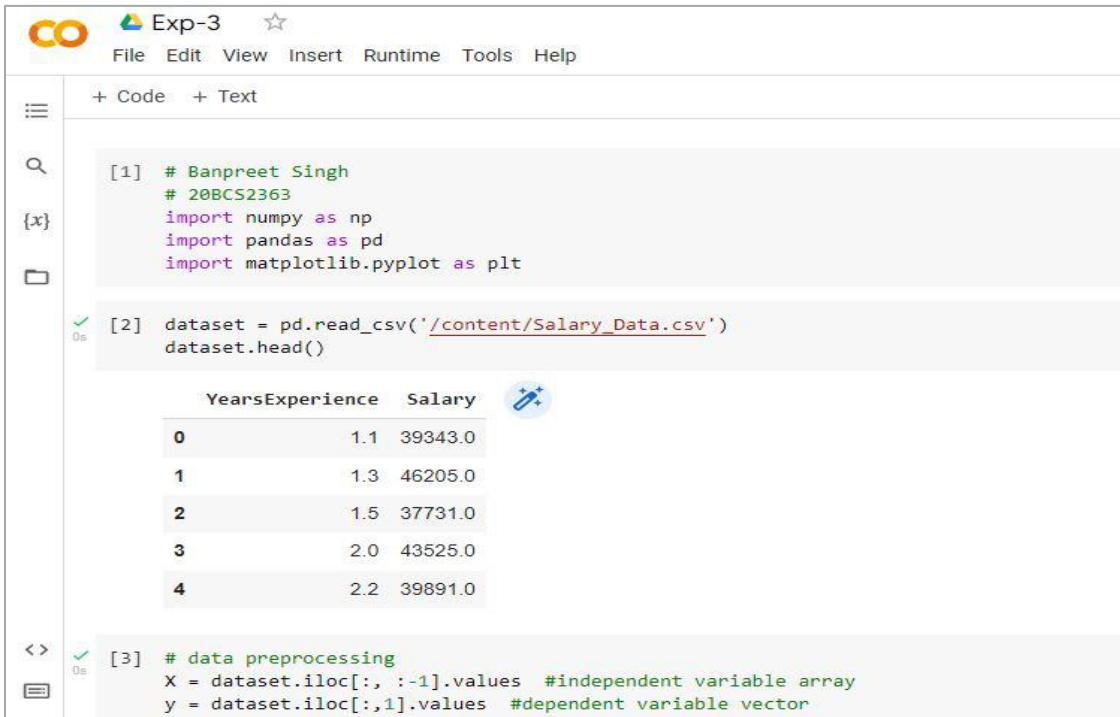
```python
# predicting the test set results
y_pred = regressor.predict(X_test)
y_pred

y_test
```

```python
#plot for the TRAIN
plt.scatter(X_train, y_train, color='red') # plotting the observation line
plt.plot(X_train, regressor.predict(X_train), color='blue') #plot the regression line
plt.title("Salary vs Experience (Training set)")
plt.xlabel("Years of experience")
plt.ylabel("Salaries")
plt.show()
```

```python
#plot for the TEST
plt.scatter(X_test, y_test, color='red')
plt.plot(X_train, regressor.predict(X_train), color='blue') #plot the regression line
plt.title("Salary vs Experience (Testing set)")
plt.xlabel("Years of experience")
plt.ylabel("Salaries")
plt.show()
```

4. Result/Output:

```
CO   Exp-3   ☆
     File  Edit  View  Insert  Runtime  Tools  Help

     + Code   + Text

[1]  # Banpreet Singh
     # 20BCS2363
     import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt

[2]  dataset = pd.read_csv('/content/Salary_Data.csv')
     dataset.head()

         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

[3]  # data preprocessing
     X = dataset.iloc[:, :-1].values   #independent variable array
     y = dataset.iloc[:,1].values   #dependent variable vector
```

CO ☁ Exp-3 ☆

File   Edit   View   Insert   Runtime   Tools   Help   All changes saved

+ Code   + Text

```
[4]  # splitting the dataset
     from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=1/3,random_state=0)
```

```
[5]  # fitting the regression model
     from sklearn.linear_model import LinearRegression
     regressor = LinearRegression()
     regressor.fit(X_train,y_train) #actually produces the linear eqn for the data

     LinearRegression()
```
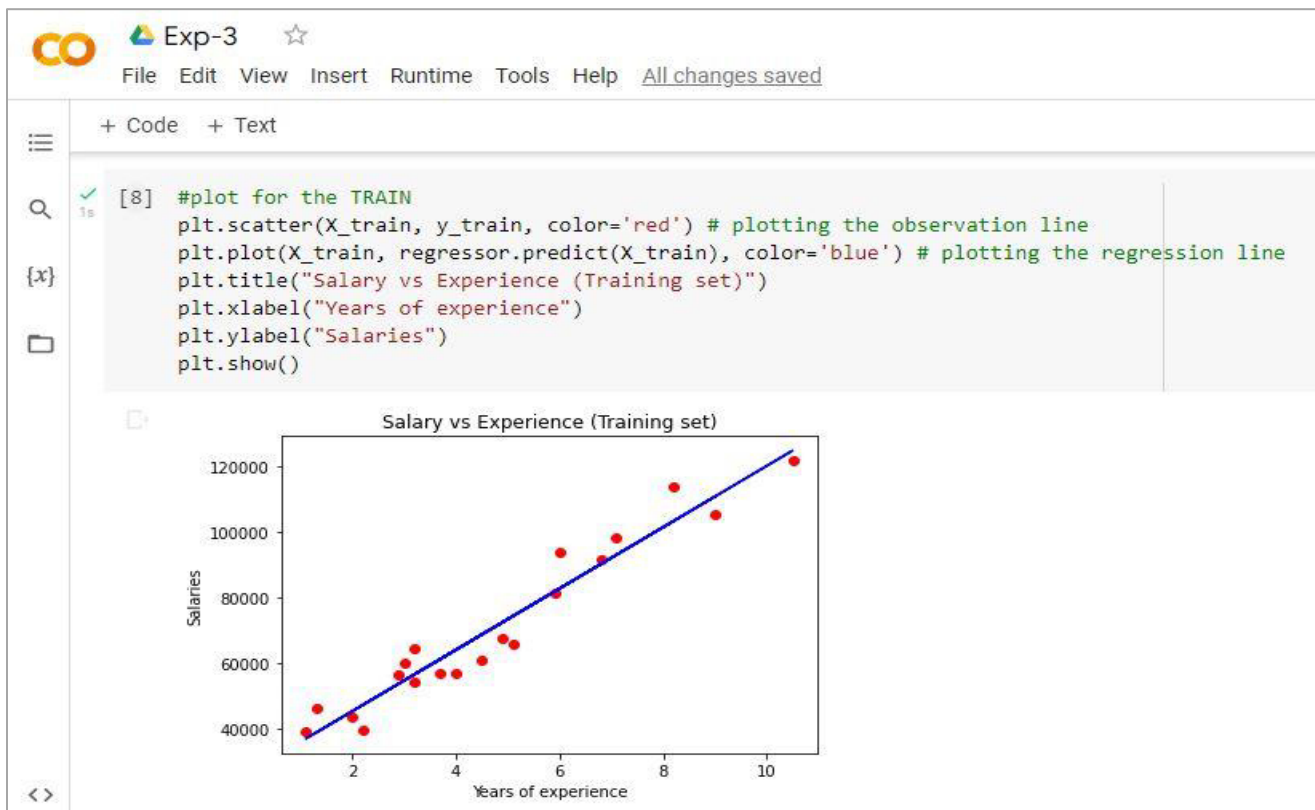
```
[6]  # predicting the test set results
     y_pred = regressor.predict(X_test)
     y_pred

     array([ 40835.10590871, 123079.39940819,  65134.55626083,  63265.36777221,
            115602.64545369, 108125.8914992 , 116537.23969801,  64199.96201652,
             76349.68719258, 100649.1375447 ])
```
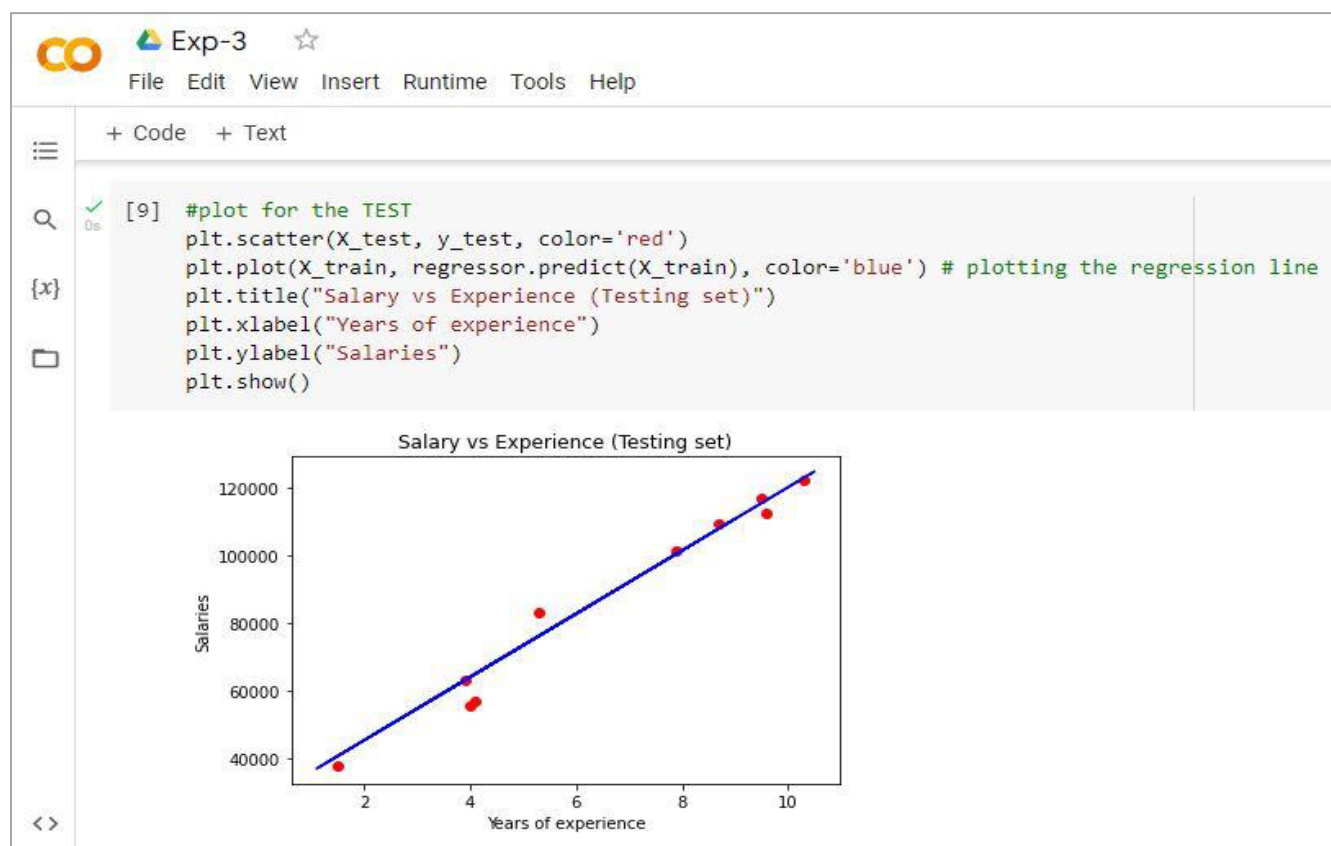
```
[7]  y_test

     array([ 37731., 122391.,  57081.,  63218., 116969., 109431., 112635.,
             55794.,  83088., 101302.])
```

CO ☁ Exp-3 ☆

File   Edit   View   Insert   Runtime   Tools   Help   All changes saved

+ Code   + Text

```
[8]  #plot for the TRAIN
     plt.scatter(X_train, y_train, color='red') # plotting the observation line
     plt.plot(X_train, regressor.predict(X_train), color='blue') # plotting the regression line
     plt.title("Salary vs Experience (Training set)")
     plt.xlabel("Years of experience")
     plt.ylabel("Salaries")
     plt.show()
```

CO Exp-3 ☆

File Edit View Insert Runtime Tools Help

+ Code + Text

[9] #plot for the TEST

```python
#plot for the TEST
plt.scatter(X_test, y_test, color='red')
plt.plot(X_train, regressor.predict(X_train), color='blue') # plotting the regression line
plt.title("Salary vs Experience (Testing set)")
plt.xlabel("Years of experience")
plt.ylabel("Salaries")
plt.show()
```


Salary vs Experience (Testing set)

5. Learning outcomes (What I have learnt):

1. Understanding of the linear regression.
2. Able to analyze different datasets with the help of python and pandas library.
3. Learning about numpy and matplotlib libraries of python.
4. Learn about the different methods/functions that are needed to perform linear regression.
5. Learning of different Machine Learning Functions

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			