PROG 8 LINEAR REGRESSION

AIM:

To write a python program to compute linear regression curve between salary and experience from a csv dataset.

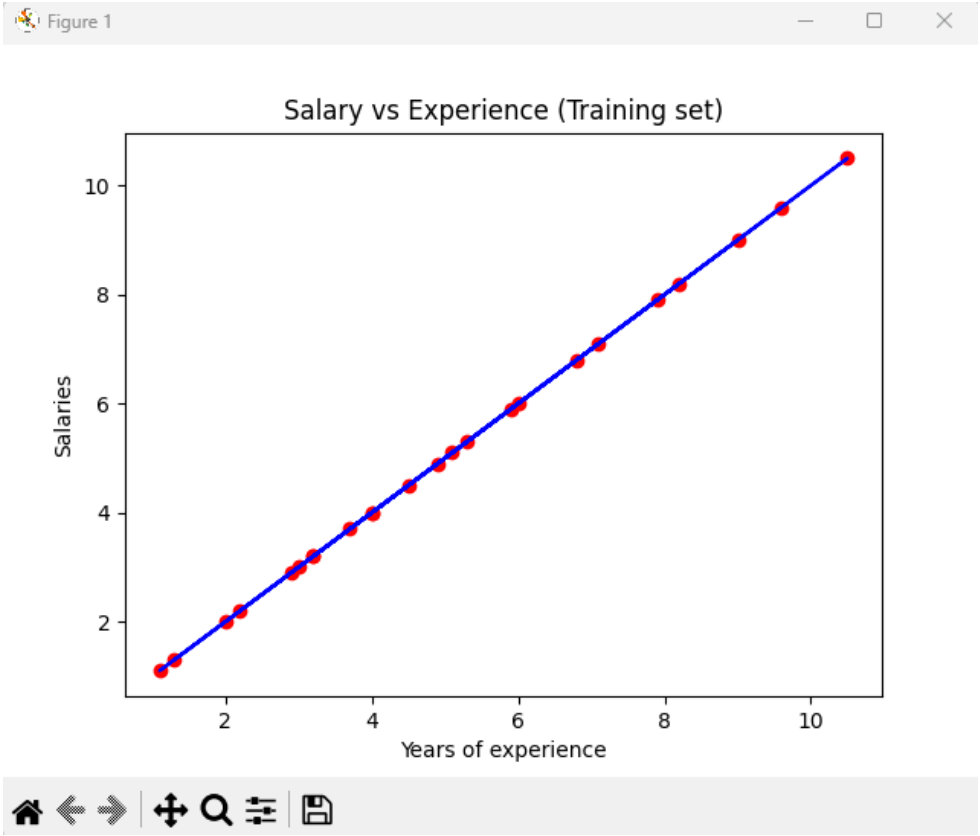
SOURCE CODE:

```
 import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

dataset = pd.read_csv("./Salary_Data.csv")
print(dataset.head())
print(dataset.info())

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=0.2,random_state=0)
#print('Training Data\n',X_train)
#print('Testing Data\n',X_test)

# 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

# Plotting the graph for the Training dataset

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)") # stating the title of the graph

plt.xlabel("Years of experience") # adding the name of x-axis
plt.ylabel("Salaries") # adding the name of y-axis
plt.show() # specifies end of graph

# Plotting the graph for the Testing dataset

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()
plt.show()
```

OUTPUT:

```
   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
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     float64
dtypes: float64(2)
memory usage: 608.0 bytes
None
```



RESULT:

Thus the program is executed and the output is verified successfully.