

# LINEAR REGRESSION

DEVELOPED BY: VISHRUTH KHARE

ROLL NUMBER: 2K18/CO/393

IDE: PyCharm

Compiler: Terminal (Python 3.6)

DataSet: Salary Dataset

Sample Dataset:

	YearsExperience	Salary
1		
2	1.1	39343.00
3	1.3	46205.00
4	1.5	37731.00
5	2.0	43525.00
6	2.2	39891.00
7	2.9	56642.00
8	3.0	60150.00
9	3.2	54445.00
10	3.2	64445.00
11	3.7	57189.00
12	3.9	63218.00
13	4.0	55794.00
14	4.0	56957.00
15	4.1	57081.00
16	4.5	61111.00
17	4.9	67938.00
18	5.1	66029.00
19	5.3	83088.00
20	5.9	81363.00
21	6.0	93940.00
22	6.8	91738.00
23	7.1	98273.00
24	7.9	101302.00
25	8.2	113812.00

Source Code:

```
# Importing the libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Importing the dataset
dataset = pd.read_csv('Salary_Data.csv')
```

```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values

# Splitting the dataset into the Training set and Test set
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 Simple Linear Regression to the Training set
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)

# Predicting the Test set results
y_pred = regressor.predict(X_test)

# Visualising the Training set results
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title("Salary vs Experience (Training set)")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()

# Visualising the Test set results
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title("Salary vs Experience (Test set)")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()
```

OUTPUT:

