

Program 4: Develop a program to implement Simple Linear Regression model and evaluate the model by verifying the performance

```
import pandas as pd
df = pd.read_csv("Salary_Data.csv")
df.head()
```

```
YearsExperience Salary
0      1.1  39343
1      1.3  46205
2      1.5  37731
3      2.0  43525
4      2.2  39891
```

```
df.shape
(30, 2)
```

Independent and Dependent Variables

```
X=df.drop('Salary',axis=1)
y=df.Salary
X.head()
```

```
YearsExperience
0      1.1
1      1.3
2      1.5
3      2.0
4      2.2
```

```
y.head()
0  39343
1  46205
2  37731
3  43525
4  39891
```

Splitting the data

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,random_state=0,test_size=0.30)
X_test.shape
(9, 1)
```

Model Fitting

```
from sklearn.linear_model import LinearRegression
LR=LinearRegression()
LR.fit(X_train,y_train)
LinearRegression()
```

Prediction

```
y_pred=LR.predict(X_test)
y_pred
```

```
array([ 40817.78327049, 123188.08258899, 65154.46261459, 63282.41035735,  
       115699.87356004, 108211.66453108, 116635.89968866, 64218.43648597,  
       76386.77615802])
```

y_test

```
2    37731  
28   122391  
13   57081  
10   63218  
26   116969  
24   109431  
27   112635  
11   55794  
17   83088  
Name: Salary
```

Evaluation

```
from sklearn import metrics  
R2=metrics.r2_score(y_test,y_pred)  
R2
```

```
0.9740993407213511
```

```
print(metrics.mean_absolute_error(y_test,y_pred))
```

```
3737.4178618788987
```

```
print(LR.predict([[5]]))
```

```
[73578.69777217]
```

Visualization

```
import seaborn as sns  
import matplotlib.pyplot as plt  
sns.regplot(X,y)  
plt.show()
```

