EXPERIMENT NO: 7

Linear Regression

Aim:

To write a Python program to understand and perform salary prediction using Linear Regression on the given dataset.

Algorithm:

1. Load the dataset and inspect its structure and summary statistics.
2. Separate the independent (Years of Experience) and dependent (Salary) variables.
3. Split the dataset into training and testing sets.
4. Train the Linear Regression model using the training data.
5. Evaluate the model performance using accuracy scores.
6. Predict salary for a given experience and save the model using Pickle.

Program:

[1]:
```
import numpy as np
import pandas as pd
df = pd.read_csv("C:/Users/siddesh/Downloads/Salary_data.csv")
df.head()
```

[1]:

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

[2]:
```
df.info()
```

```
<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     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
```

[3]:
```
df.dropna(inplace=True)
```

[4]:
```
df.info()
```

```
<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     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
```

[5]:
```
df.describe()
```

[5]:

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

[6]:
```
features = df.iloc[:, [0]].values
label = df.iloc[:, [1]].values
```

[7]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(features, label, test_size=0.2, random_state=0)
```

[8]:
```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)
```

[8]:
```
LinearRegression
LinearRegression()
```

[9]:
```
model.score(x_train, y_train)
```

[9]:
```
0.9411949620562126
```

[10]:
```
model.score(x_test, y_test)
```

[10]:
```
0.988169515729126
```

[11]:
```
model.coef_
```

[11]:
```
array([[9312.57512673]])
```

[12]:
```
model.intercept_
```

[12]:
```
array([26780.09915063])
```

[13]:
```
import pickle
pickle.dump(model, open('SalaryPred.model', 'wb'))
```

[14]:
```
model = pickle.load(open('SalaryPred.model', 'rb'))
```

[16]:
```
yr_of_exp = 10
yr_of_exp_NP = np.array([[yr_of_exp]])
Salary = model.predict(yr_of_exp_NP)
print(f"Estimated Salary for {yr_of_exp} years of experience is ₹{Salary[0][0]:,.2f}")
```

```
Estimated Salary for 10 years of experience is ₹119,905.85
```

Result:

Thus, the Python program is executed successfully for predicting salary from the dataset using Linear Regression.