

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

```
In [4]: data = pd.read_csv("Salary_Data_31f066cc745250401d28bfebe00bad27.csv")
data
```

Out[4]:

	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
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0
10	3.9	63218.0
11	4.0	55794.0
12	4.0	56957.0
13	4.1	57081.0
14	4.5	61111.0
15	4.9	67938.0
16	5.1	66029.0
17	5.3	83088.0
18	5.9	81363.0
19	6.0	93940.0
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

```
In [5]: data.head(10)
```

Out[5]:

	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
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0

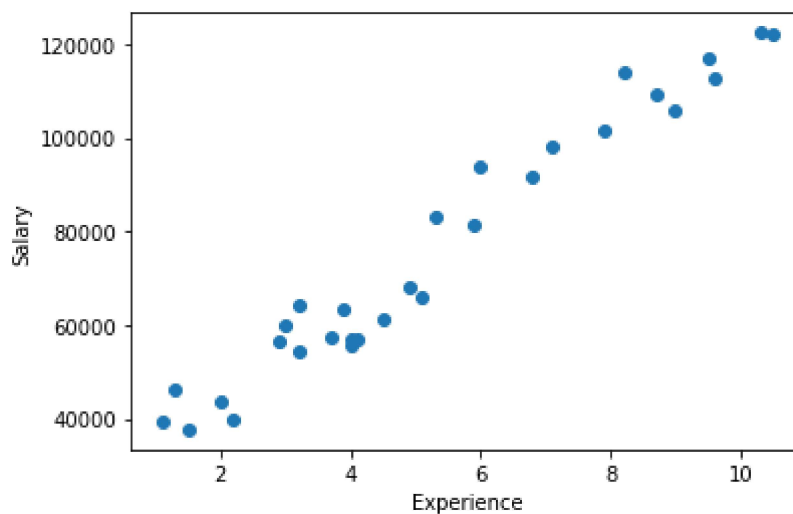
```
In [6]: x = np.array(data["YearsExperience"]).reshape(-1,1)
x
```

```
Out[6]: array([[ 1.1],
 [ 1.3],
 [ 1.5],
 [ 2. ],
 [ 2.2],
 [ 2.9],
 [ 3. ],
 [ 3.2],
 [ 3.2],
 [ 3.7],
 [ 3.9],
 [ 4. ],
 [ 4. ],
 [ 4.1],
 [ 4.5],
 [ 4.9],
 [ 5.1],
 [ 5.3],
 [ 5.9],
 [ 6. ],
 [ 6.8],
 [ 7.1],
 [ 7.9],
 [ 8.2],
 [ 8.7],
 [ 9. ],
 [ 9.5],
 [ 9.6],
 [10.3],
 [10.5]])
```

```
In [7]: y = data["Salary"]  
y
```

```
Out[7]: 0      39343.0  
1      46205.0  
2      37731.0  
3      43525.0  
4      39891.0  
5      56642.0  
6      60150.0  
7      54445.0  
8      64445.0  
9      57189.0  
10     63218.0  
11     55794.0  
12     56957.0  
13     57081.0  
14     61111.0  
15     67938.0  
16     66029.0  
17     83088.0  
18     81363.0  
19     93940.0  
20     91738.0  
21     98273.0  
22    101302.0  
23    113812.0  
24    109431.0  
25    105582.0  
26    116969.0  
27    112635.0  
28    122391.0  
29    121872.0  
Name: Salary, dtype: float64
```

```
In [8]: plt.scatter(x,y)  
plt.xlabel("Experience")  
plt.ylabel("Salary")  
plt.show()
```



```
In [9]: # Data splitting  
X_train, X_test, y_train, y_test = train_test_split(x,y,test_size = 0.3)  
  
print(len(X_train))  
print(len(X_test))
```

21

9

```
In [10]: model = LinearRegression()  
model.fit(X_train,y_train)
```

```
Out[10]: LinearRegression()
```

```
In [11]: i = model.predict([[4]])  
print(i)
```

[62694.36226424]

```
In [12]: # Evaluate the model  
acc = model.score(X_test,y_test)  
print(acc)
```

0.9421236678648653

```
In [ ]:
```