

```
In [1]: import pandas as pd
df=pd.read_csv('Salary_Data.csv')
print(df)
```

	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 [2]: print("Shape of data:",df.shape)
```

```
Shape of data: (30, 2)
```

```
In [3]: print("Missing Data: \n",df.isna().sum())
```

```
Missing Data:  
YearsExperience    0  
Salary            0  
dtype: int64
```

```
In [4]: print(df.describe())
```

	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

```
In [5]: x=df['YearsExperience'].values.reshape(-1,1)  
y=df['Salary'].values.reshape(-1,1)  
print(x)
```

```
[[ 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 [6]: `print(y)`

```
[[ 39343.]  
 [ 46205.]  
 [ 37731.]  
 [ 43525.]  
 [ 39891.]  
 [ 56642.]  
 [ 60150.]  
 [ 54445.]  
 [ 64445.]  
 [ 57189.]  
 [ 63218.]  
 [ 55794.]  
 [ 56957.]  
 [ 57081.]  
 [ 61111.]  
 [ 67938.]  
 [ 66029.]  
 [ 83088.]
```

```
[ 81363.]  
[ 93940.]  
[ 91738.]  
[ 98273.]  
[101302.]  
[113812.]  
[109431.]  
[105582.]  
[116969.]  
[112635.]  
[122391.]  
[121872.]]
```

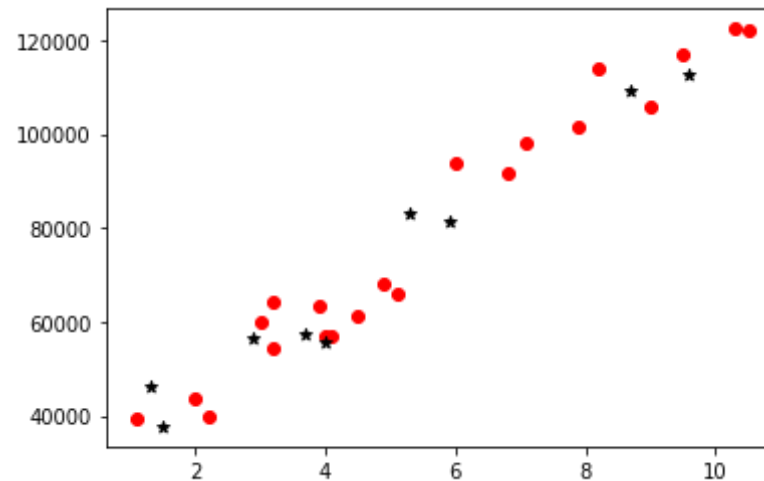
```
In [9]: import matplotlib.pyplot as plt  
plt.title("Salary vs experience")  
plt.xlabel("Experience")  
plt.ylabel("Salary")  
plt.scatter(x,y, color='b')  
plt.show()
```



```
In [10]: from sklearn.model_selection import train_test_split
```

```
x_train,x_test, y_train,y_test=train_test_split(x,y, test_size=0.3, random_state=7)
plt.scatter(x_train,y_train, color='r', marker='o')
plt.scatter(x_test, y_test, color='k', marker='*')
```

Out[10]: <matplotlib.collections.PathCollection at 0x21e1e5b2ac8>



```
In [11]: from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(x_train,y_train)
```

Out[11]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

```
In [12]: print("Accuracy: ",model.score(x_test,y_test))
```

Accuracy: 0.958816457461513

```
In [13]: y_pred=model.predict(x)
plt.title("Salary vs experience")
plt.xlabel("Experience")
plt.ylabel("Salary")
plt.scatter(x,y, color='b', label="Actual")
```

```
plt.plot(x,y_pred, color='r',label="pred")  
plt.legend()  
plt.show()
```



```
In [14]: model.predict([[12]])
```

```
Out[14]: array([[140230.00199725]])
```

```
In [ ]:
```