regression_and_exercise_7

November 2, 2025

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

[2]:
```
df.head(5)
```

[2]:

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

[3]:
```
df.dropna()
```

[3]:

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891
5	2.9	56642
6	3.0	60150
7	3.2	54445
8	3.2	64445
9	3.7	57189
10	3.9	63218
11	4.0	55794
12	4.0	56957
13	4.1	57081
14	4.5	61111
15	4.9	67938
16	5.1	66029
17	5.3	83088
18	5.9	81363
19	6.0	93940
20	6.8	91738
21	7.1	98273
22	7.9	101302

```
23              8.2  113812
24              8.7  109431
25              9.0  105582
26              9.5  116969
27              9.6  112635
28             10.3  122391
29             10.5  121872
```

[4]:
```
x=df.iloc[:,[0]].values
y=df.iloc[:,[1]].values
```

[5]:
```
from sklearn.model_selection import train_test_split
```

[6]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

[7]:
```
from sklearn.linear_model import LinearRegression
```

[8]:
```
model=LinearRegression()#this is the stage where i create a model which has no␣
↪data an empty model with no knowledge
```

[9]:
```
model
```

[9]:
```
LinearRegression()
```

[10]:
```
model.fit(x_train,y_train)#model is trained with the data of x and y
```

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

[11]:
```
model.predict([[5]])
```

[11]:
```
array([[73342.97478427]])
```

[12]:
```
y_pred=model.predict(x_test)
```

[13]:
```
y_pred
```

[13]:
```
array([[ 40748.96184072],
       [122699.62295594],
       [ 64961.65717022],
       [ 63099.14214487],
       [115249.56285456],
       [107799.50275317]])
```

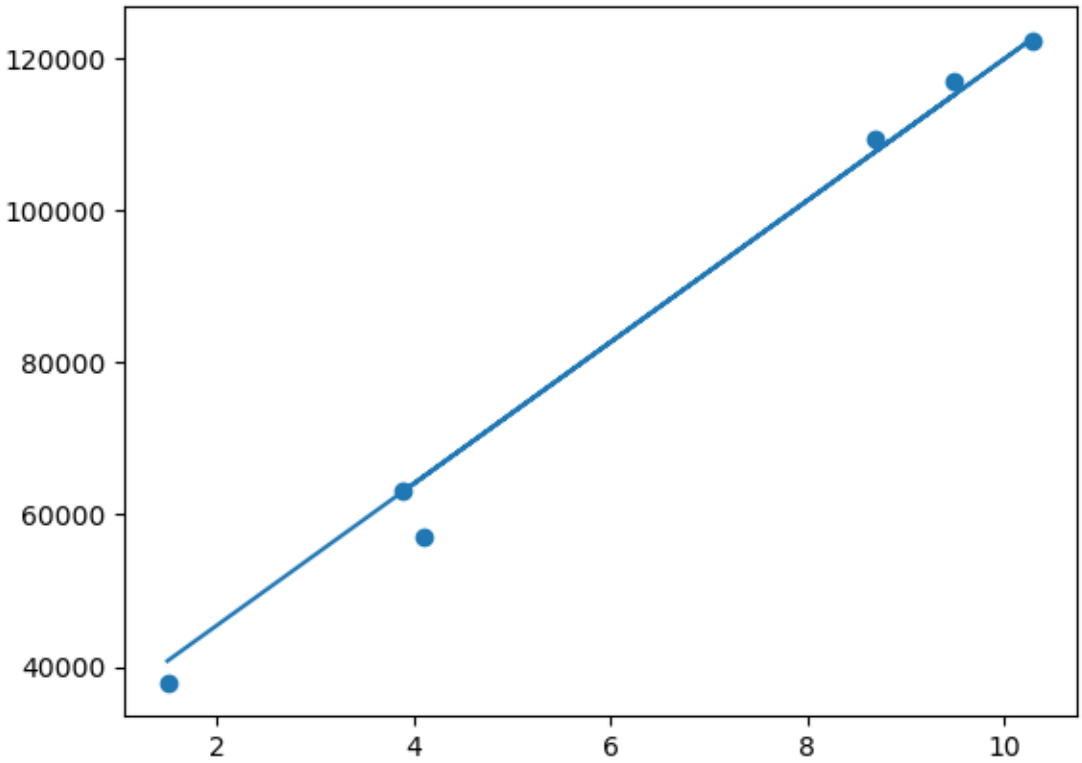
[14]:
```
errors=y_pred-y_test
errors
```

[14]:
```
array([[ 3017.96184072],
       [  308.62295594],
```

```
       [ 7880.65717022],
       [ -118.85785513],
       [-1719.43714544],
       [-1631.49724683]])
```

[15]:
```
import matplotlib.pyplot as plt
plt.scatter(x_test,y_test)
plt.plot(x_test,y_pred)
```

[15]: [<matplotlib.lines.Line2D at 0x26ea0de2fd0>]



[16]:
```
from sklearn.metrics import r2_score
accuracy=r2_score(y_test,y_pred)
```

[17]:
```
accuracy
```

[17]: 0.988169515729126

[18]:
```
model.predict([[44]])
```

[18]: array([[436533.40472671]])

[19]:
```
model.score(x_train,y_train)#This tells how the model regression fits this model
```

[19]:
```
0.9411949620562126
```

[20]:
```
model.score(x_test,y_test)
```

[20]:
```
0.988169515729126
```

[21]:
```
model.coef_#the coefficient is the slope of the best-fit line.
```

[21]:
```
array([[9312.57512673]])
```

[22]:
```
model.intercept_
```

[22]:
```
array([26780.09915063])
```

[23]:
```
model.predict([[55]])
```

[23]:
```
array([[538971.73112073]])
```

[ ]:
```

```