linear-regression-1

April 15, 2025

[1]:
```
# Linear Regression
```

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
dataset = pd.read_csv("/content/Salary_Data.csv")
```

[4]:
```
dataset
```

[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
```

[5]:
```
x = dataset.iloc[:,0:1].values
```

[6]:
```
x
```

[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]])
```

[7]:
```
y = dataset.iloc[:,-1].values
```

[8]:
```
y
```

[8]:
```
array([ 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.])
```

[9]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.
 ↪2,random_state=10)
```

[10]:
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train,y_train)
```

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

[11]:
```
regressor.coef_
```

[11]:
```
array([9356.86299354])
```

[12]:
```
regressor.intercept_
```

[12]:
```
np.float64(26089.09663241673)
```

[13]:
```
#salary = 9356*Exp + 26089
```

[14]:
```
x_new = pd.read_csv('/content/Salary_Data.csv')
```

[15]:
```
x_new = x_new.iloc[:,:].values
```

[16]:
```
x_new
```

[16]:
```
array([[1.10000e+00, 3.93430e+04],
       [1.30000e+00, 4.62050e+04],
       [1.50000e+00, 3.77310e+04],
       [2.00000e+00, 4.35250e+04],
       [2.20000e+00, 3.98910e+04],
       [2.90000e+00, 5.66420e+04],
       [3.00000e+00, 6.01500e+04],
       [3.20000e+00, 5.44450e+04],
       [3.20000e+00, 6.44450e+04],
       [3.70000e+00, 5.71890e+04],
       [3.90000e+00, 6.32180e+04],
       [4.00000e+00, 5.57940e+04],
       [4.00000e+00, 5.69570e+04],
       [4.10000e+00, 5.70810e+04],
       [4.50000e+00, 6.11110e+04],
       [4.90000e+00, 6.79380e+04],
       [5.10000e+00, 6.60290e+04],
```

```
       [5.30000e+00, 8.30880e+04],
       [5.90000e+00, 8.13630e+04],
       [6.00000e+00, 9.39400e+04],
       [6.80000e+00, 9.17380e+04],
       [7.10000e+00, 9.82730e+04],
       [7.90000e+00, 1.01302e+05],
       [8.20000e+00, 1.13812e+05],
       [8.70000e+00, 1.09431e+05],
       [9.00000e+00, 1.05582e+05],
       [9.50000e+00, 1.16969e+05],
       [9.60000e+00, 1.12635e+05],
       [1.03000e+01, 1.22391e+05],
       [1.05000e+01, 1.21872e+05]])
```

[17]:
```
y_pred_test=regressor.predict(x_test)
```

[18]:
```
x_new = pd.read_csv('/content/Salary_Data.csv')
# Select only the 'YearsExperience' column for prediction, similar to how x was␣
 ↪defined during training.
x_new = x_new.iloc[:, 0:1].values
y_pred_new = regressor.predict(x_new)
```

[19]:
```
y_pred_test
```

[19]:
```
array([89715.76498848, 56031.05821174, 53223.99931368, 40124.39112273,
       44802.8226195 , 92522.82388655])
```

[20]:
```
y_pred_new
```

[20]:
```
array([ 36381.64592531,  38253.01852402,  40124.39112273,  44802.8226195 ,
        46674.1952182 ,  53223.99931368,  54159.68561303,  56031.05821174,
        56031.05821174,  60709.48970851,  62580.86230722,  63516.54860657,
        63516.54860657,  64452.23490593,  68194.98010334,  71937.72530076,
        73809.09789947,  75680.47049818,  81294.5882943 ,  82230.27459365,
        89715.76498848,  92522.82388655, 100008.31428138, 102815.37317944,
       107493.80467621, 110300.86357427, 114979.29507104, 115914.98137039,
       122464.78546587, 124336.15806458])
```

[21]:
```
from sklearn.metrics import r2_score
r2_score(y_test, y_pred_test)
```

[21]:
```
0.9816423482070253
```

[22]:
```
y_test
```

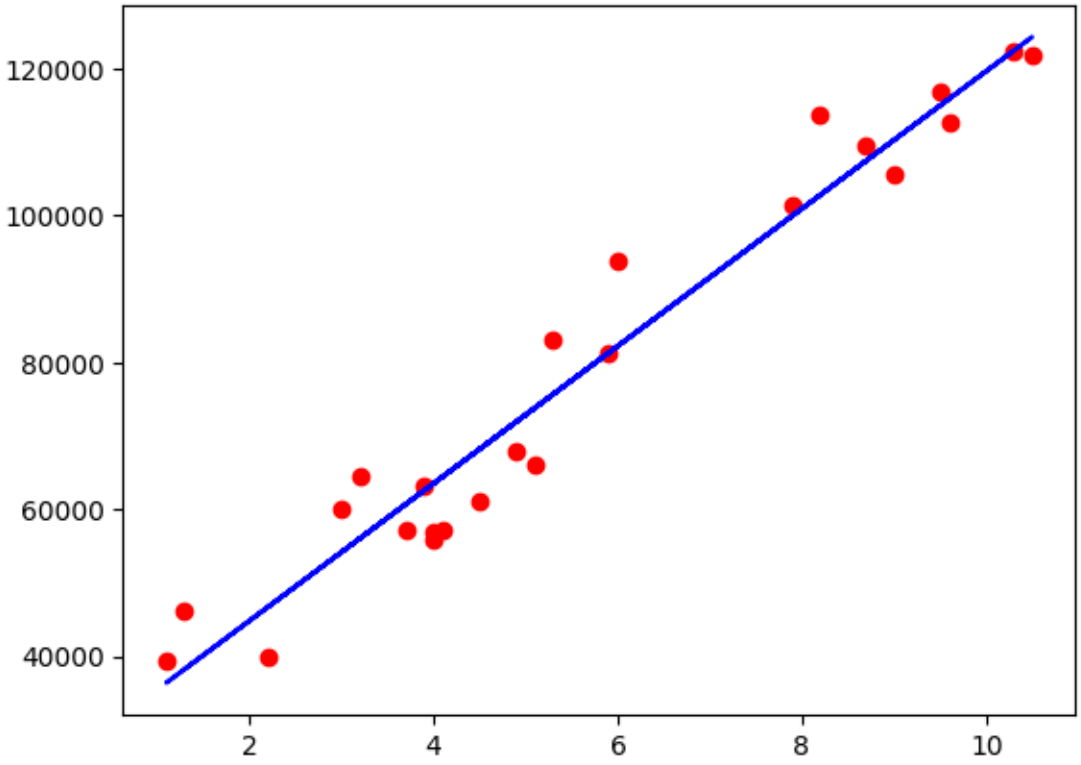
[22]:
```
array([91738., 54445., 56642., 37731., 43525., 98273.])
```

[ ]:
```
y_pred_new
```

[ ]:
```
array([ 36381.64592531,  38253.01852402,  40124.39112273,  44802.8226195 ,
        46674.1952182 ,  53223.99931368,  54159.68561303,  56031.05821174,
        56031.05821174,  60709.48970851,  62580.86230722,  63516.54860657,
        63516.54860657,  64452.23490593,  68194.98010334,  71937.72530076,
        73809.09789947,  75680.47049818,  81294.5882943 ,  82230.27459365,
        89715.76498848,  92522.82388655, 100008.31428138, 102815.37317944,
       107493.80467621, 110300.86357427, 114979.29507104, 115914.98137039,
       122464.78546587, 124336.15806458])
```

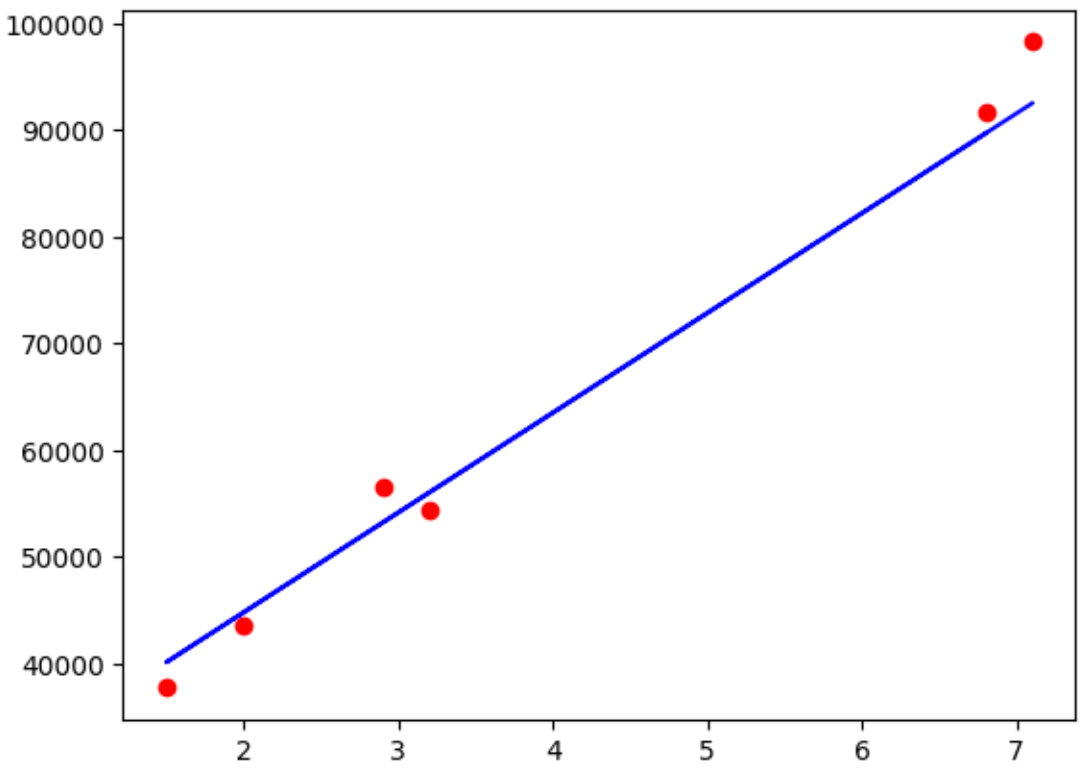
[23]:
```
plt.scatter(x_train,y_train, color = 'red')
plt.plot(x_train, regressor.predict(x_train), color='blue')
```

[23]:
```
[<matplotlib.lines.Line2D at 0x7b5973992a50>]
```



[24]:
```
plt.scatter(x_test,y_test, color = 'red')
plt.plot(x_test, regressor.predict(x_test), color='blue')
```

[24]:
```
[<matplotlib.lines.Line2D at 0x7b59728ce790>]
```


100000
90000
80000
70000
60000
50000
40000
2
3
4
5
6
7