practical03

September 18, 2024

Rohit Himmat Koli232121004

[26]:
```
import matplotlib as sns
import pandas as pd
import numpy as np

data =pd.read_csv('student_scores.csv')
```

[27]:
```
data
```

[27]:

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30
5	1.5	20
6	9.2	88
7	5.5	60
8	8.3	81
9	2.7	25
10	7.7	85
11	5.9	62
12	4.5	41
13	3.3	42
14	1.1	17
15	8.9	95
16	2.5	30
17	1.9	24
18	6.1	67
19	7.4	69
20	2.7	30
21	4.8	54
22	3.8	35
23	6.9	76
24	7.8	86

[28]:
```
data.describe
```

[28]:
```
<bound method NDFrame.describe of     Hours  Scores
0     2.5      21
1     5.1      47
2     3.2      27
3     8.5      75
4     3.5      30
5     1.5      20
6     9.2      88
7     5.5      60
8     8.3      81
9     2.7      25
10    7.7      85
11    5.9      62
12    4.5      41
13    3.3      42
14    1.1      17
15    8.9      95
16    2.5      30
17    1.9      24
18    6.1      67
19    7.4      69
20    2.7      30
21    4.8      54
22    3.8      35
23    6.9      76
24    7.8      86>
```

[29]:
```
x=data.iloc[:,:-1].values
y=data.iloc[:,1].values
```

[30]:
```
x
```

[30]:
```
array([[2.5],
       [5.1],
       [3.2],
       [8.5],
       [3.5],
       [1.5],
       [9.2],
       [5.5],
       [8.3],
       [2.7],
       [7.7],
       [5.9],
       [4.5],
       [3.3],
       [1.1],
```

```
       [8.9],
       [2.5],
       [1.9],
       [6.1],
       [7.4],
       [2.7],
       [4.8],
       [3.8],
       [6.9],
       [7.8]])
```

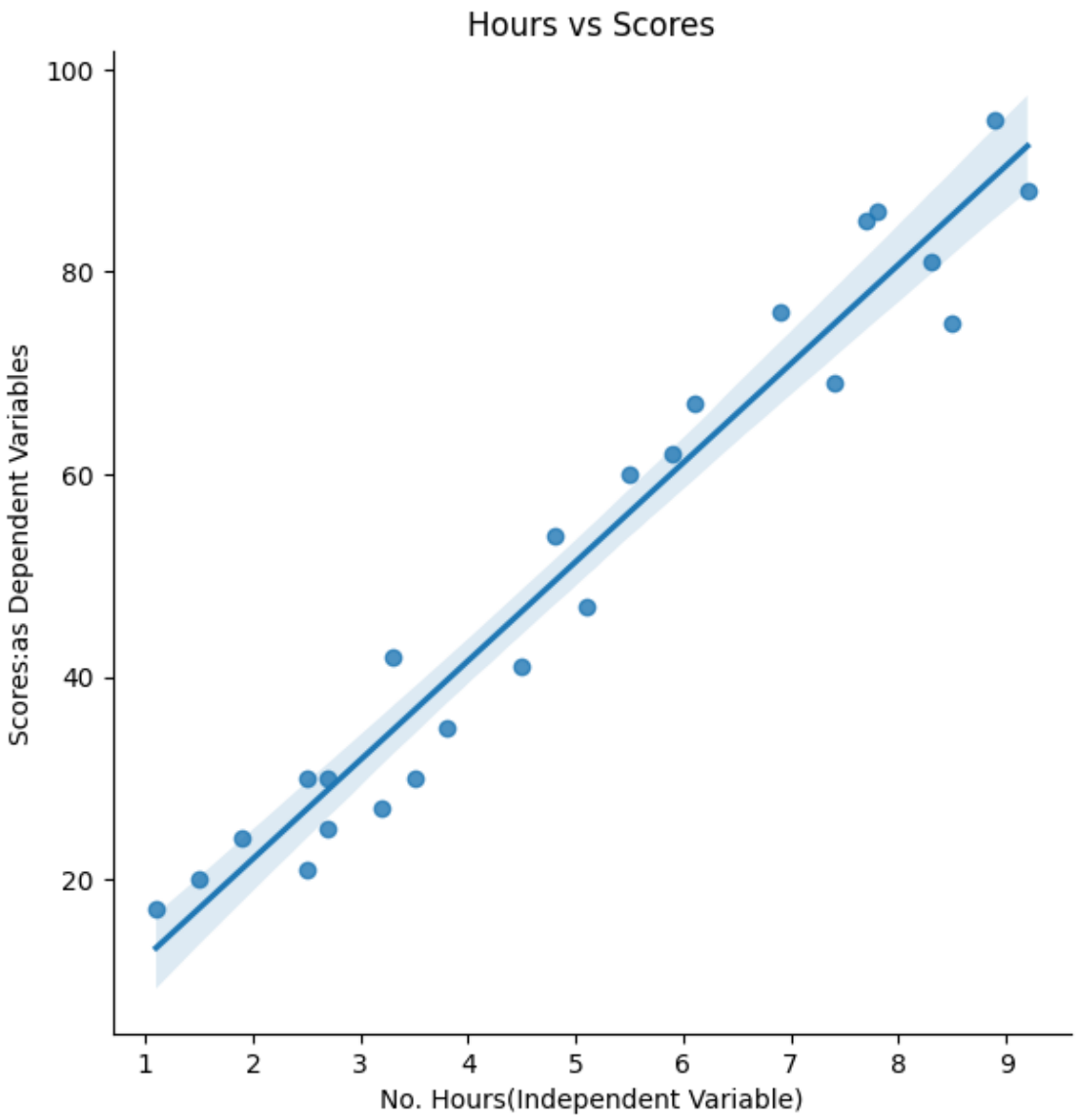
[31]:
```
y
```

[31]:
```
array([21, 47, 27, 75, 30, 20, 88, 60, 81, 25, 85, 62, 41, 42, 17, 95, 30,
       24, 67, 69, 30, 54, 35, 76, 86], dtype=int64)
```

[32]:
```
import matplotlib.pyplot as plt
import seaborn as sns

sns.lmplot(x='Hours',y='Scores',data=data,height=6)
plt.xlabel('No. Hours(Independent Variable)')
plt.ylabel('Scores:as Dependent Variables')
plt.title('Hours vs Scores')
```

[32]:
```
Text(0.5, 1.0, 'Hours vs Scores')
```



[33]:
```
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=0)
```

[34]:
```
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(x_train,y_train)
```

[34]:
```
LinearRegression()
```

[35]:
```
x_train
```

[35]:
```
array([[3.8],
       [1.9],
       [7.8],
       [6.9],
       [1.1],
       [5.1],
       [7.7],
       [3.3],
       [8.3],
       [9.2],
       [6.1],
       [3.5],
       [2.7],
       [5.5],
       [2.7],
       [8.5],
       [2.5],
       [4.8],
       [8.9],
       [4.5]])
```

[36]:
```
x_test
```

[36]:
```
array([[1.5],
       [3.2],
       [7.4],
       [2.5],
       [5.9]])
```

[37]:
```
y_pred=regressor.predict(x_test)
```

[38]:
```
y_pred
```

[38]:
```
array([16.88414476, 33.73226078, 75.357018  , 26.79480124, 60.49103328])
```

[39]:
```
regressor.predict([[2]])
```

[39]:
```
array([21.839473])
```

[40]:
```
from sklearn import metrics
print('Mean absolute error :',metrics.mean_absolute_error(y_test,y_pred))
print('Mean squad error :',metrics.mean_squared_error(y_test,y_pred))
print('Root mean squared error :',np.sqrt(metrics.
  ↪mean_squared_error(y_test,y_pred)))
```

```
Mean absolute error : 4.183859899002975
Mean squad error : 21.598769307217406
Root mean squared error : 4.647447612100367
```