ASSIGNMENT / TASK 7

Prediction using Supervised Machine Learning using Simple Linear Regression

In this task we have to find the students scores based on their study hours. This is a simple Regression problem type because it has only two variables.

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv("/content/StudentHoursScores (1).csv")
print(df)
```

```
    Hours  Scores
0    7.7      79
1    5.9      60
2    4.5      45
3    3.3      33
4    1.1      12
5    8.9      87
6    2.5      21
7    1.9      19
8    2.7      29
9    8.3      81
10   5.5      58
11   9.2      88
12   1.5      14
13   3.5      34
14   8.5      85
15   3.2      32
16   6.5      66
17   2.5      21
18   9.6      96
19   4.3      42
20   4.1      40
21   3.0      30
22   2.6      25
```

```python
df.head()
```

	Hours	Scores
0	7.7	79
1	5.9	60

```
df.shape
```

```
(23, 2)
```

```
df.describe()
```

	Hours	Scores
count	23.000000	23.000000
mean	4.817391	47.695652
std	2.709688	27.103228
min	1.100000	12.000000
25%	2.650000	27.000000
50%	4.100000	40.000000
75%	7.100000	72.500000
max	9.600000	96.000000

```
import sklearn
from sklearn.model_selection import train_test_split
x=df.iloc[:,:-1].values
y=df.iloc[:,1].values
xtrain,xtest,ytrain,ytest=train_test_split(x,y, test_size=1/3, random_state=1)
from sklearn.linear_model import LinearRegression
model=LinearRegression()
mymodel=model.fit(xtrain,ytrain)
y_pred=mymodel.predict(xtest)
print("predicted value\n",y_pred)
print("Value of b\n",mymodel.intercept_)
print("Value of a\n",mymodel.coef_)
plt.scatter(xtrain,ytrain,color='red')
plt.plot(xtrain,mymodel.predict(xtrain), label="line of regression")#mymodel.predict(xtrain):
plt.legend()
plt.xlabel("HOURS")
plt.ylabel("SCORES")
plt.show()
```
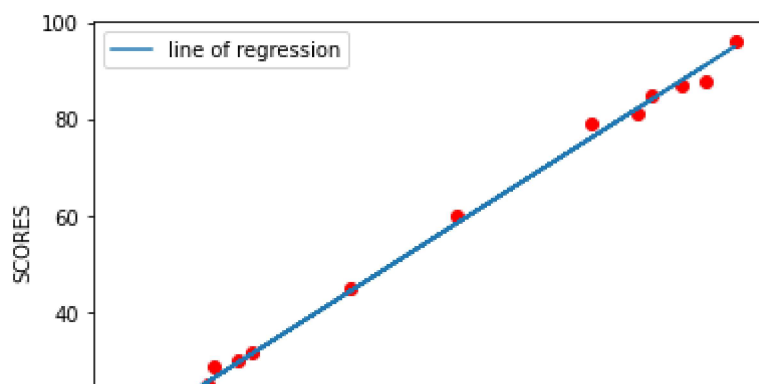
```
predicted value
 [40.50038393 24.63061172 32.56549782 34.54921935 42.48410545 64.30504223
 54.3864346  10.74456104]
Value of b
 -0.16590734578961275
Value of a
 [9.91860763]
```



```python
model.predict([[5]])
```

```
array([49.42713079])
```