THE SPARK FOUNDATION GRIP

TASK1:PREDICTION USING SUPERVISED MACHINE LEARNING

AUTHOR: KAVIYA S

In [11]:

```
  #required packages
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [12]:

```
#reading the data
data=pd.read_csv("C:/Users/kaviya subramanian/Desktop/DATA.csv")
data.head()
```

Out[12]:

	hours	scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

In [13]:

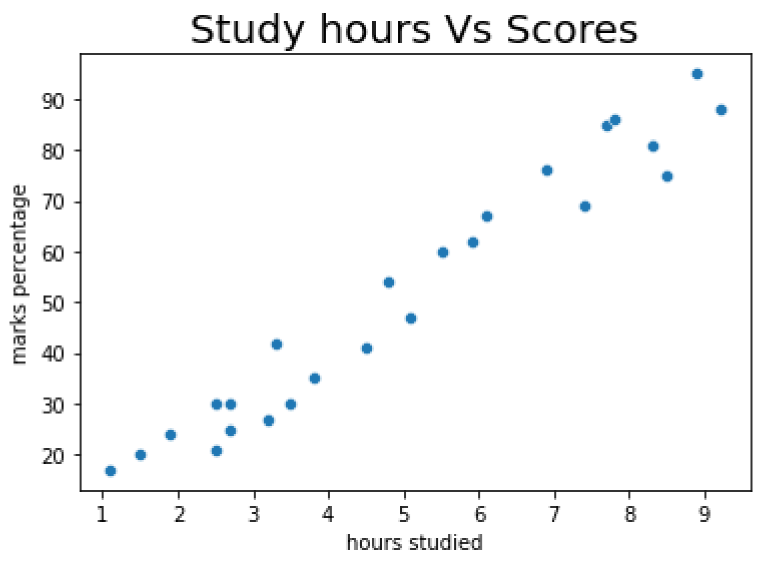
```
#chech null value
data.isnull==True
```

Out[13]:

```
False
```

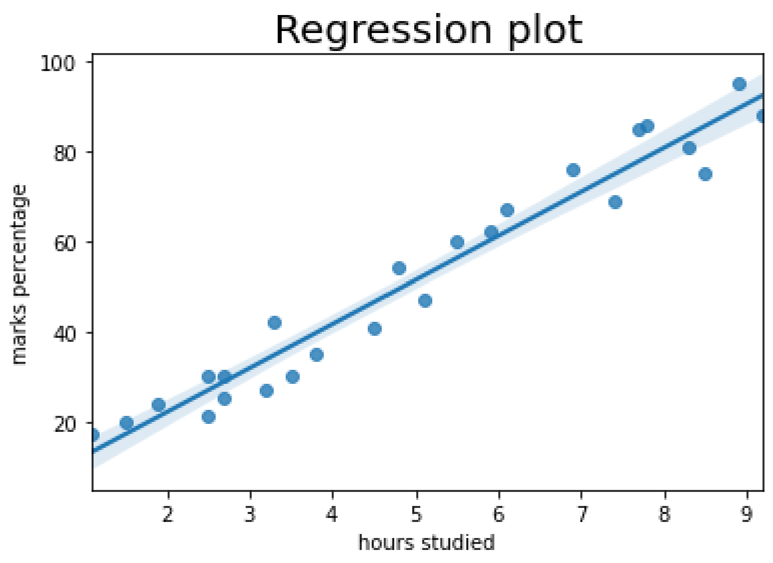
In [14]:

```
sns.scatterplot(y=data["scores"],x=data["hours"])
plt.title("Study hours Vs Scores",size=20)
plt.ylabel("marks percentage")
plt.xlabel("hours studied")
plt.show()
```



In [15]:

```
sns.regplot(x=data["hours"],y=data["scores"])
plt.title("Regression plot",size=20)
plt.ylabel("marks percentage")
plt.xlabel("hours studied")
plt.show()
print(data.corr())
```



```
           hours    scores
hours   1.000000  0.976191
scores  0.976191  1.000000
```

In [16]:

```
#definig x and y from the data
x=data.iloc[:,:-1].values
y=data.iloc[:,1].values
```

In [17]:

```
#spliting the data in two
train_x,test_x,train_y,test_y=train_test_split(x,y,random_state=0)
```

In [18]:

```
regression=LinearRegression()
regression.fit(train_x,train_y)
print("model trained")
```

```
model trained
```

In [19]:

```
pred_y=regression.predict(test_x)
prediction=pd.DataFrame({"hours":[i[0] for i in test_x],"predicted marks":[k for k in pred_y]})
prediction
```

Out[19]:

	hours	predicted marks
0	1.5	16.844722
1	3.2	33.745575
2	7.4	75.500624
3	2.5	26.786400
4	5.9	60.588106
5	3.8	39.710582
6	1.9	20.821393

In [20]:

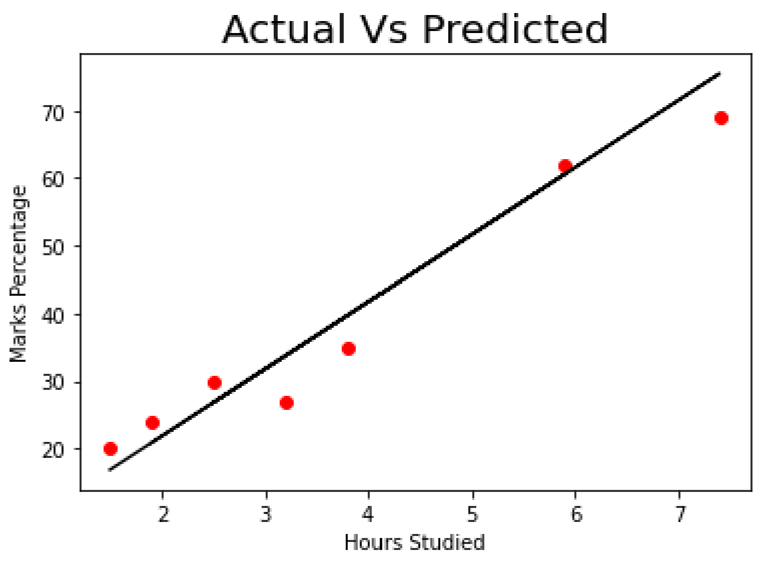
```
compare_scores=pd.DataFrame({"Actual Marks":test_y,"Predicted Marks":pred_y})
compare_scores
```

Out[20]:

	Actual Marks	Predicted Marks
0	20	16.844722
1	27	33.745575
2	69	75.500624
3	30	26.786400
4	62	60.588106
5	35	39.710582
6	24	20.821393

In [21]:

```
plt.scatter(x=test_x,y=test_y,color="red")
plt.plot(test_x,pred_y,color="black")
plt.title("Actual Vs Predicted",size=20)
plt.ylabel("Marks Percentage")
plt.xlabel("Hours Studied")
plt.show()
```



In [22]:

```
#calculating the accuracy of the model
from sklearn.metrics import mean_absolute_error
print("mean absolute error:",mean_absolute_error(test_y,pred_y))
```

```
mean absolute error: 4.130879918502482
```

In [23]:

```
hours=[9.25]
answer=regression.predict([hours])
print("scores={}".format(round(answer[0],3)))
```

```
scores=93.893
```

CONCLUSION

from the above result we can say that if a studied for 9.25 then student will secured 93.893 MARKS.