NAME : Mahalakshmi S

TASK 1 : Prediction using Supervised ML

Predict the percentage of an student based on the no. of study hours.

# GRIPJAN21

# DATASCIENCE

In [1]:
```
#Importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
#Import data
url="http://bit.ly/w-data"
data=pd.read_csv(url)
data.head(25)
```

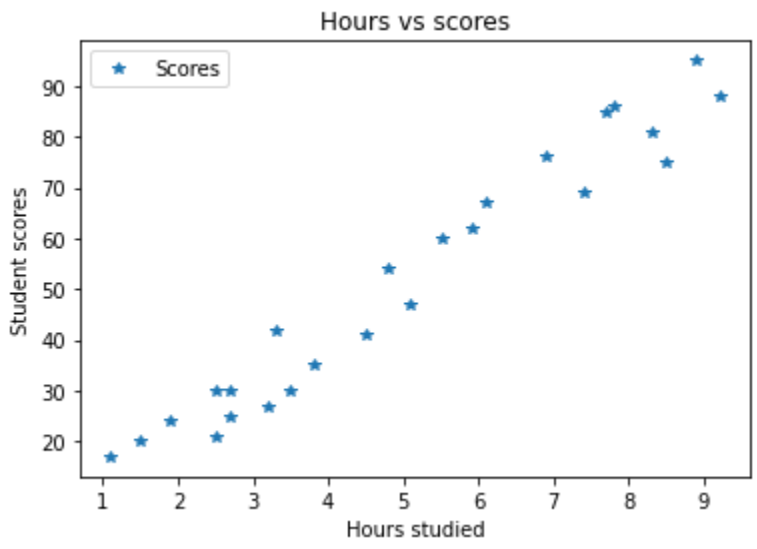
Out[2]:

| | 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 |

In [3]:
```
data.shape
```

Out[3]: (25, 2)

In [8]:
```
#Plot the data
data.plot(x='Hours',y='Scores',style='*')
plt.title('Hours vs scores')
plt.xlabel('Hours studied')
plt.ylabel('Student scores')
plt.show()
```



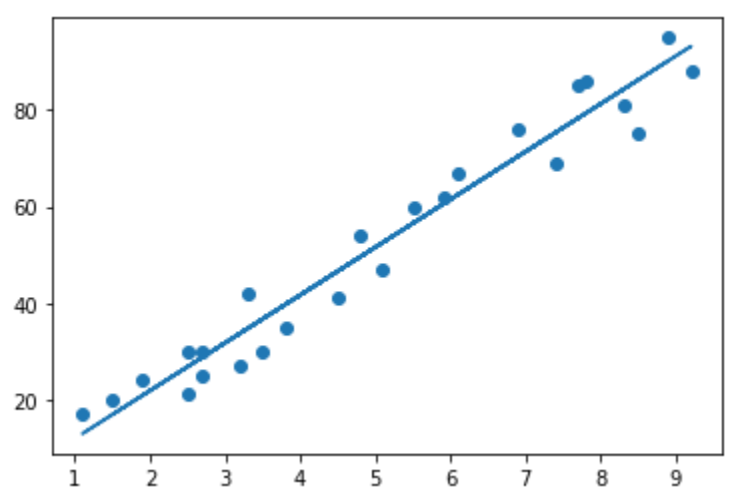

In [9]:
```
#Splitting variables
x=data.iloc[:,:-1].values
y=data.iloc[:,1].values
```

In [10]:
```
#train_test split
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)
```

In [17]:
```
from sklearn.linear_model import LinearRegression
regr=LinearRegression()
regr.fit(x_train,y_train)
print('Training completed')
```

```
Training completed
```

In [18]:
```
#Plotting the line
fit_line=regr.coef_*x + regr.intercept_
plt.scatter(x,y)
plt.plot(x,fit_line)
plt.show()
```




In [19]:
```
#Testing
print(x_test)
```

```
[[1.5]
 [3.2]
 [7.4]
 [2.5]
 [5.9]]
```

In [20]:
```
#Predicting scores
y_pred=regr.predict(x_test)
```

In [22]:
```
df=pd.DataFrame({"Actual":y_test,"Predicted":y_pred})
df
```

Out[22]:

| | Actual | Predicted |
|---|---|---|
| **0** | 20 | 16.884145 |
| **1** | 27 | 33.732261 |
| **2** | 69 | 75.357018 |
| **3** | 30 | 26.794801 |
| **4** | 62 | 60.491033 |

In [23]:
```
#Prediction for 9.25 hours
hours=9.25
my_pred=regr.predict([[hours]])
print("No of hours studied{}".format(hours))
print("Predicted score{}".format(my_pred[0]))
```

```
No of hours studied9.25
Predicted score93.69173248737538
```

In [ ]:
```
#Thankyou
```