A simple Linear Regression model algorithm that predicts the score of a student based on the number of hours they studied.

DataSet : http://bit.ly/w-data

In [2]:

```
#Importing all the required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score,mean_squared_error,mean_absolute_error
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [3]:

```
#reading the data
url='http://bit.ly/w-data'
df=pd.read_csv(url)
df.head()
```

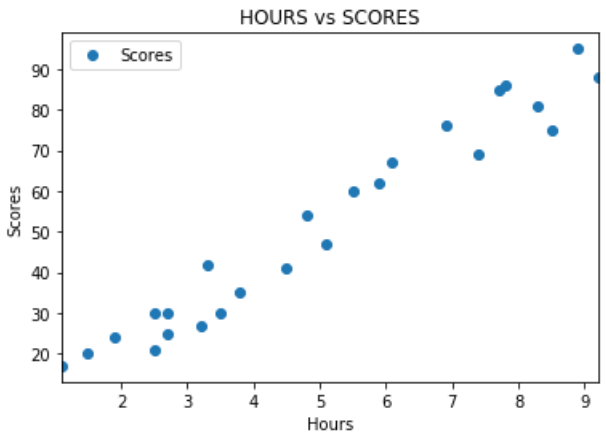
Out[3]:

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

In [4]:

```
#plotting the data
plt.figure(figsize=(6,4))
df.plot('Hours','Scores',style='o')
plt.title('HOURS vs SCORES')
plt.xlabel('Hours')
plt.ylabel('Scores')
plt.show()
```

```
<Figure size 432x288 with 0 Axes>
```



In [5]:

```
x=np.array(df['Hours']).reshape(-1,1)
y=np.array(df['Scores']).reshape(-1, 1)
```

In [6]:

```
#splitting the data for training and testing
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
x_train.shape
```

Out[6]:

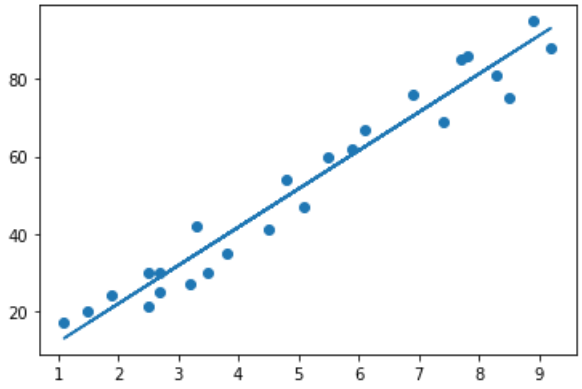
```
(20, 1)
```

In [7]:

```
#Fitting the data
regressor=LinearRegression().fit(x_train,y_train)
```

In [8]:

```
line=regressor.coef_*x+regressor.intercept_
#plotting for the test data
plt.scatter(x,y)
#plotting the regression line
plt.plot(x,line)
plt.show()
```



In [9]:

```
y_pred=regressor.predict(x_test)
print(f'predicted score is {y_pred}')
```

```
predicted score is [[16.88414476]
 [33.73226078]
 [75.357018  ]
 [26.79480124]
 [60.49103328]]
```

In [10]:

```
#prediction for 9.25 hours
y_pred1=regressor.predict([[9.25]])
print(f'no of hours is {9.25}')
print(f'predicted score is {y_pred1[0]}')
```

```
no of hours is 9.25
predicted score is [93.69173249]
```

In [11]:

```
#Evaluation metrics
print(f' mean squared error :{mean_squared_error(y_test,y_pred)}')
print(f' mean absolute error:{mean_absolute_error(y_test,y_pred)}')
print(f' R2 score           : {r2_score(y_test,y_pred)}')
```

```
 mean squared error :21.598769307217456
 mean absolute error:4.183859899002982
```

```
Mean absolute error:4.183859899002982
R2 score            : 0.9454906892105354
```