```
In [1]:  import numpy as np
         import matplotlib.pyplot as plt
         import pandas as pd
         import seaborn as sns
```

```
In [6]:  url="http://bit.ly/w-data"
         dataset=pd.read_csv(url)
         x=dataset.iloc[:,:-1].values
         y=dataset.iloc[:,-1].values
```

```
In [7]:  x
```

```
Out[7]:  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]])
```

In [8]: `y`

Out[8]:
```
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)
```

In [9]: `dataset`

Out[9]:

	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

	Hours	Scores
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 [10]: `dataset.describe()`

Out[10]:

	Hours	Scores
count	25.000000	25.000000
mean	5.012000	51.480000
std	2.525094	25.286887
min	1.100000	17.000000
25%	2.700000	30.000000
50%	4.800000	47.000000
75%	7.400000	75.000000
max	9.200000	95.000000

In [12]: `dataset.shape`

Out[12]: `(25, 2)`

In [13]: `dataset.head(5)`

Out[13]:

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

In [14]: `dataset.tail()`

Out[14]:

	Hours	Scores
20	2.7	30
21	4.8	54
22	3.8	35
23	6.9	76
24	7.8	86

In [15]:
```python
plt.scatter(x,y)
plt.title("scatter plot of hours vs score")
plt.xlabel("Hours")
plt.ylabel("Scores")
```

Out[15]: `Text(0, 0.5, 'Scores')`

scatter plot of hours vs score
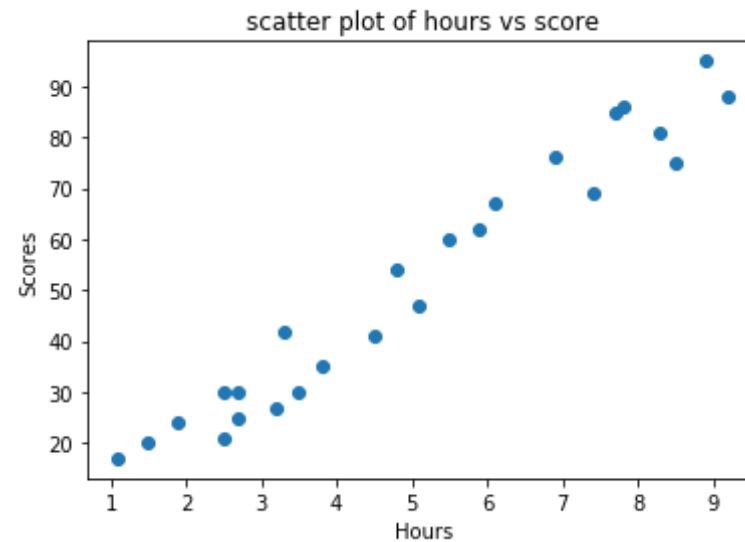
```
In [16]:   dataset.isnull().sum()

Out[16]:   Hours     0
           Scores    0
           dtype: int64
```

```
In [42]:   from sklearn.model_selection import train_test_split
           x_train,x_test,y_train,y_test= train_test_split(x,y,test_size=0.3,rando
           m_state=0)
```

```
In [43]:   from sklearn.linear_model import LinearRegression
           regressor=LinearRegression()
           regressor.fit(x_train,y_train)

Out[43]:   LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normaliz
           e=False)
```

```
In [44]:   plt.scatter(x,y)
           plt.plot(x,regressor.coef_*x+regressor.intercept_,color='g')
           plt.xlabel('Hours studied')
```

```
plt.ylabel('Percentage Score')
plt.show()
```
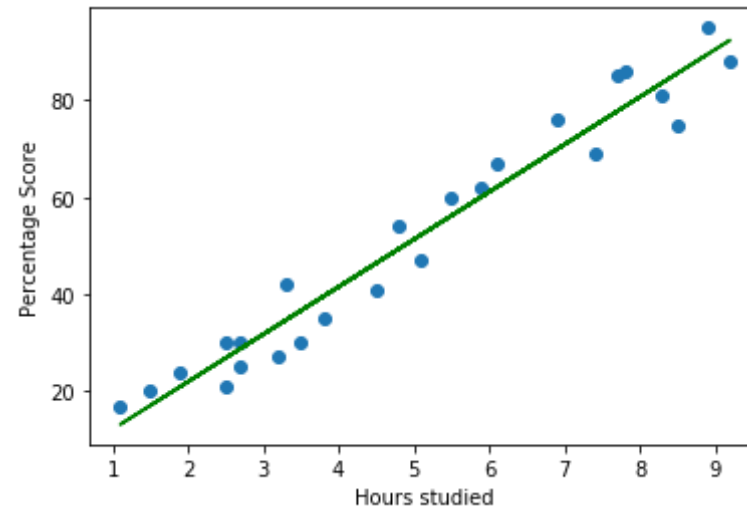

In [45]: `prediction=regressor.predict(x_test)`

In [46]: `prediction`

Out[46]: `array([17.05366541, 33.69422878, 74.80620886, 26.8422321 , 60.12335883,`
        `       39.56736879, 20.96909209, 78.72163554])`

In [50]: `df=pd.DataFrame({'Actual':y_test,'Predicted':prediction})`

In [48]: `df`

Out[48]:

	Actual	Predicted
0	20	17.053665
1	27	33.694229
2	69	74.806209
3	30	26.842232

	Actual	Predicted
4	62	60.123359
5	35	39.567369
6	24	20.969092
7	86	78.721636

In [51]:
```python
prediction_score=regressor.predict([[10]])
Hours=10
print('Hours:',Hours)
print('Predicted score is:',prediction_score)
```

```
Hours: 10
Predicted score is: [100.25648225]
```

In [52]:
```python
from sklearn import metrics
print("Mean Absolute Error",metrics.mean_absolute_error(y_test,prediction))
```

```
Mean Absolute Error 4.419727808027652
```

In [ ]:

In [ ]: