In [8]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [9]:

```
dataset = pd.read_csv('student_scores.csv')
```

In [10]:

```
dataset.head()
```

Out[10]:

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

In [11]:

```
X = dataset.iloc[:,:-1].values
X
y = dataset.iloc[:,1].values
y
```

Out[11]:

```
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])
```

In [15]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=1/3,random_state=0)
```

In [16]:

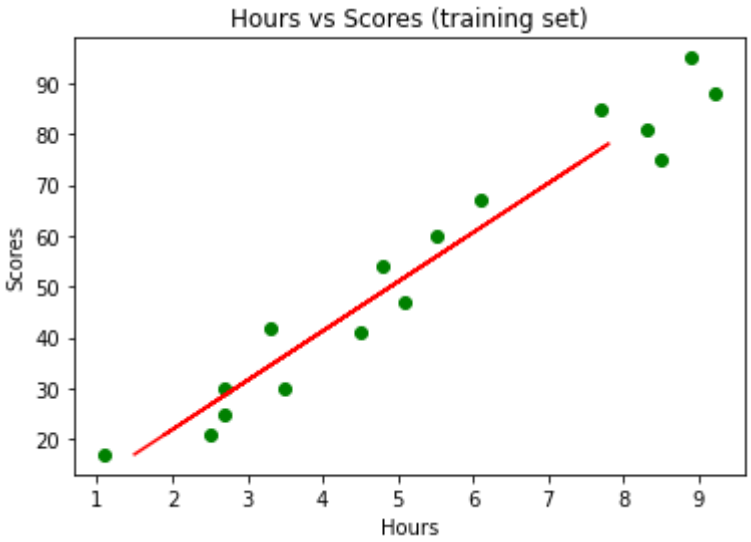
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train,y_train)
y_pred = regressor.predict(X_test)
y_pred
y_test
```

Out[16]:

```
array([20, 27, 69, 30, 62, 35, 24, 86, 76])
```

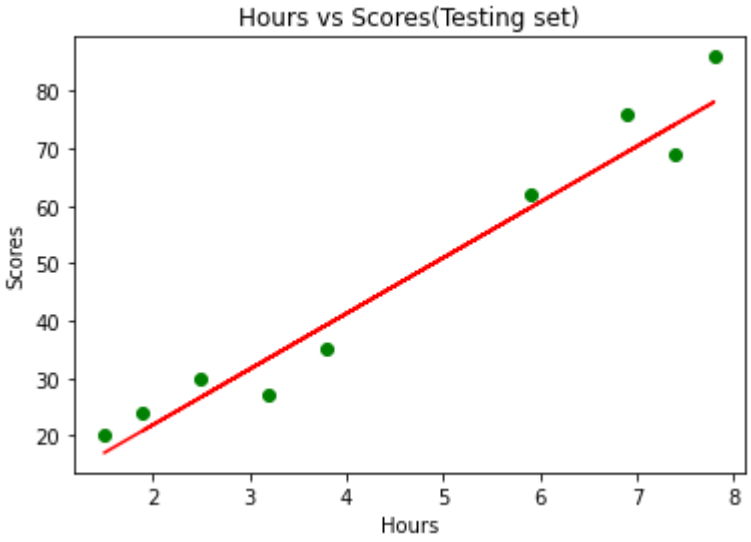
In [19]:

```
plt.scatter(X_train,y_train,color='green')
plt.plot(X_test,regressor.predict(X_test),color='red')
plt.title("Hours vs Scores (training set)")
plt.xlabel("Hours")
plt.ylabel("Scores")
plt.show()
```



In [20]:

```
plt.scatter(X_test,y_test,color='green')
plt.plot(X_test,regressor.predict(X_test),color='red')
plt.title("Hours vs Scores(Testing set)")
plt.xlabel("Hours")
plt.ylabel("Scores")
plt.show()
```



In [ ]: