

In [1]:

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

In [2]:

```
dataset = pd.read_csv('./GPA_SAT_data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
print(X)
print(y)
```

```
[[2.714]
 [3.418]
 [4.325]
 [2.552]
 [2.634]
 [3.352]
 [2.758]
 [4.231]
 [2.975]
 [3.839]
 [3.788]
 [2.133]
 [1.136]
 [3.053]
 [4.074]
 [3.515]
 [1.98 ]
 [2.788]
 [3.681]
 [4.141]
 [3.13 ]
 [2.311]
 [2.421]
 [1.631]
 [3.472]
 [1.37 ]
 [2.498]
 [3.89 ]
 [3.952]
 [2.254]
 [3.193]
 [2.835]
 [3.287]
 [3.587]]
[1480 1620 2070 1200 1300 1310 880 2230 1180 1900 1480 720 760 1490
 1940 1400 1170 1410 1290 2390 1410 1090 1120 940 1460 980 1460 1880
 2250 1480 1400 1390 1350 1850]
```

In [3]:

```
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 [4]:

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

Out[4]:

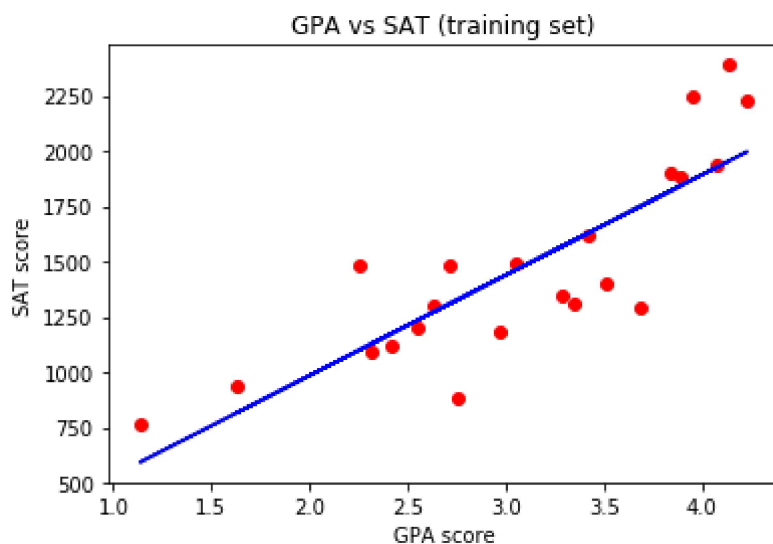
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)
```

In [7]:

```
y_pred = regressor.predict(X_test)
```

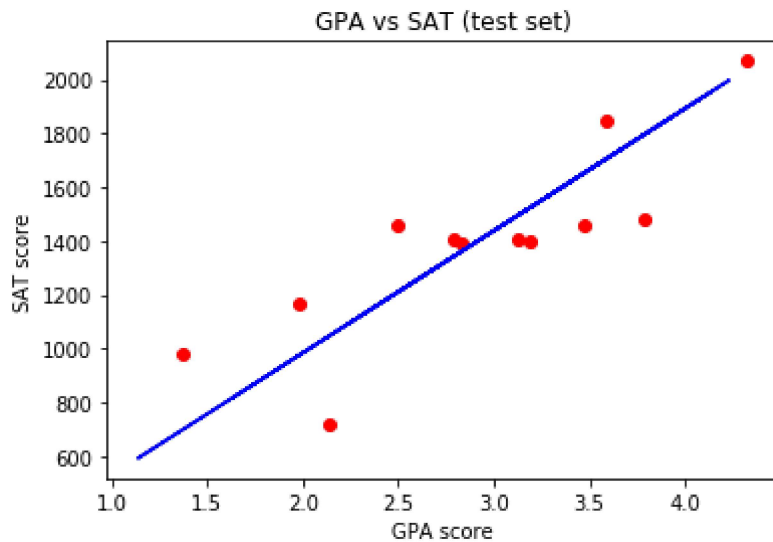
In [10]:

```
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('GPA vs SAT (training set)')
plt.xlabel('GPA score')
plt.ylabel('SAT score')
plt.show()
```



In [11]:

```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('GPA vs SAT (test set)')
plt.xlabel('GPA score')
plt.ylabel('SAT score')
plt.show()
```



In [14]:

```
GPA=float(input("enter the GPA score"))
a= regressor.predict(GPA)

print(int(a[0]))
```

```
enter the GPA score4.9
2301
```