In [1]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from seaborn import load_dataset
from sklearn.linear_model import LogisticRegression,LinearRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.model_selection import train_test_split
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
df = load_dataset('iris')
df.head()
```

Out[2]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

In [4]:

```
df['species'].unique()
```

Out[4]:

```
array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

In [5]:

```
df1=df[df['species']!='versicolor']
df1
```

Out[5]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

100 rows × 5 columns

In [6]:

```
lb = LabelEncoder()
lb
```

Out[6]:

```
LabelEncoder()
```

In [7]:

```
df1['species']=lb.fit_transform(df1['species'])
df1
```

Out[7]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	1
146	6.3	2.5	5.0	1.9	1
147	6.5	3.0	5.2	2.0	1
148	6.2	3.4	5.4	2.3	1
149	5.9	3.0	5.1	1.8	1

100 rows × 5 columns

In [8]:

```
df1['species'].unique()
```

Out[8]:

```
array([0, 1])
```

In [9]:

```
lr = LinearRegression()
lr
```

Out[9]:

```
LinearRegression()
```

In [10]:

```
x=df1.iloc[:,0].values.reshape(-1,1)
x
```

Out[10]:

```
array([[5.1],
       [4.9],
       [4.7],
       [4.6],
       [5. ],
       [5.4],
       [4.6],
       [5. ],
       [4.4],
       [4.9],
       [5.4],
       [4.8],
       [4.8],
       [4.3],
       [5.8],
       [5.7],
       [5.4],
       [5.1],
```

In [11]:

```
x.shape
```

Out[11]:

```
(100, 1)
```

In [12]:

```
y = df1.iloc[:,4].values.reshape(-1,1)
y
```

Out[12]:

```
array([[0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
       [0],
```

In [13]:

```
y.shape
```

Out[13]:

```
(100, 1)
```

In [14]:

```
lr.fit(x,y)
```

Out[14]:

```
LinearRegression()
```

In [15]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

In [16]:

```
lr.fit(x_train,y_train)
```

Out[16]:

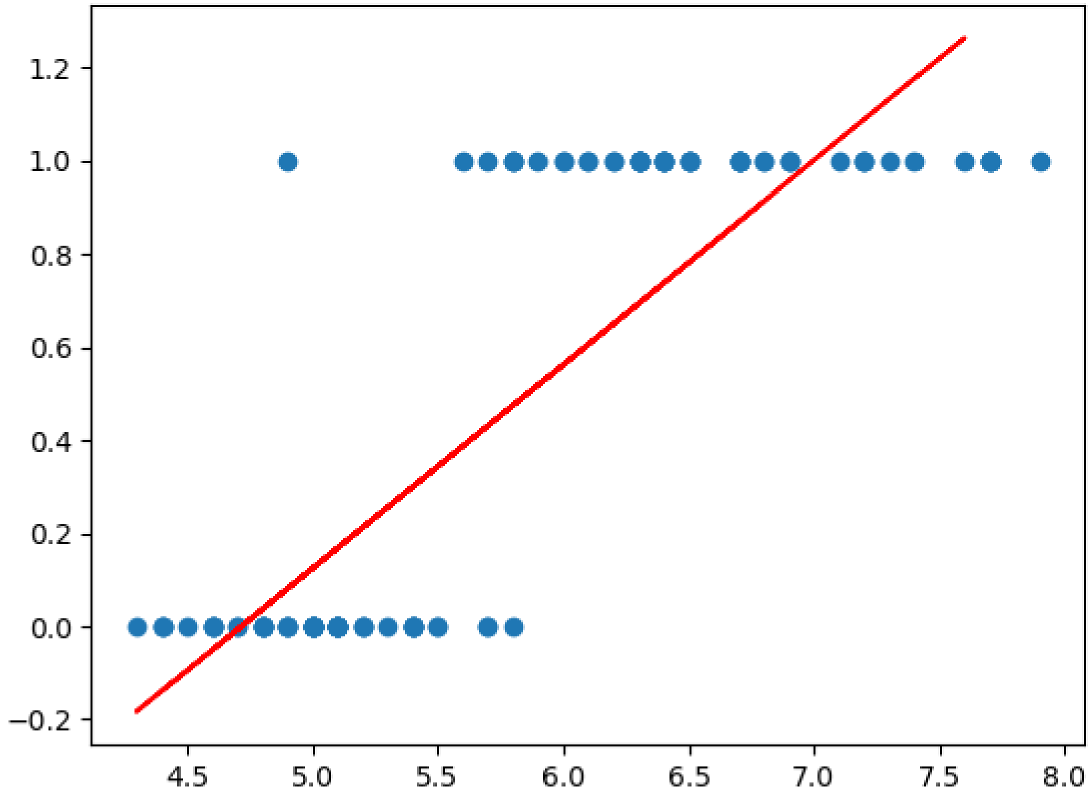
```
LinearRegression()
```

In [17]:

```
plt.scatter(x,y)
plt.plot(x_test,lr.predict(x_test),color='r')
```

Out[17]:

```
[<matplotlib.lines.Line2D at 0x222ce135dc0>]
```



In [18]:

```
loglr=LogisticRegression()
```

In [19]:

```
loglr.fit(x_test,y_test)
```

Out[19]:

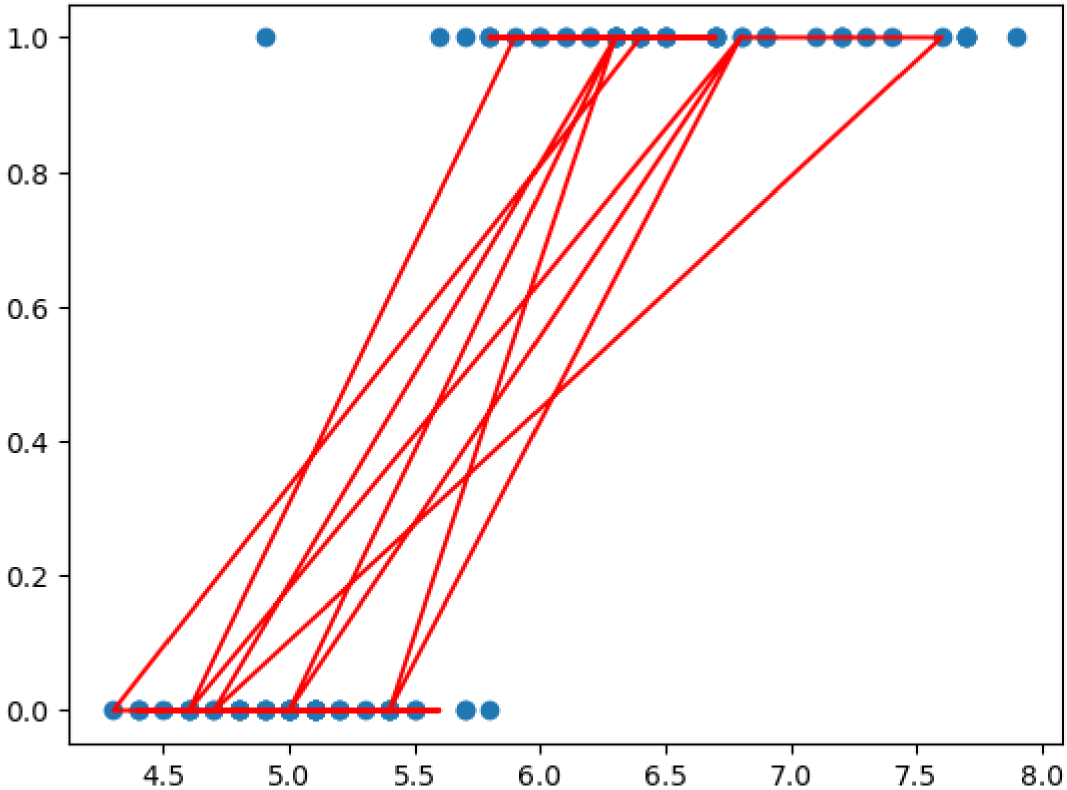
```
LogisticRegression()
```

In [20]:

```
plt.scatter(x,y)
plt.plot(x_test,loglr.predict(x_test),color='r')
```

Out[20]:

```
[<matplotlib.lines.Line2D at 0x222ce1b0640>]
```

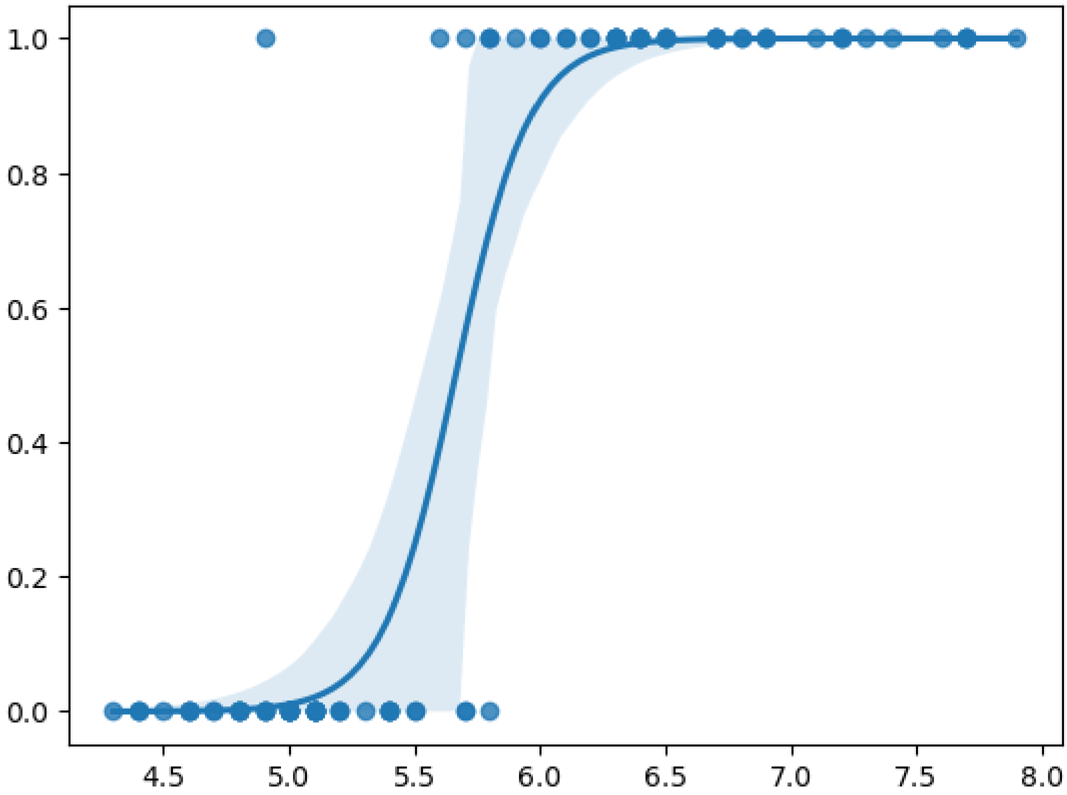


In [21]:

```
sns.regplot(x,y,logistic = True)
```

Out[21]:

```
<AxesSubplot:>
```



Confusion Matrix

In [22]:

```
y_pred = loglr.predict(x_test)

y_pred
```

Out[22]:

```
array([0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 1, 1, 1])
```

In [23]:

```
df_new1 = pd.DataFrame(y_test)

df_new2 = pd.DataFrame(y_pred)

pd.concat([df_new1, df_new2] , axis= 1)
```

Out[23]:

	0	0
0	0	0
1	1	1
2	0	0
3	1	1
4	1	1
5	1	1
6	0	0
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	1	1
22	0	0
23	1	0
24	0	0
25	0	0
26	0	0
27	1	1
28	1	1
29	1	1

In [24]:

```
confusion_matrix(y_test, y_pred)
```

Out[24]:

```
array([[15,  0],
       [ 1, 14]], dtype=int64)
```

In [25]:

```
(15 + 14)/(15 + 0 + 1 + 14)
```

Out[25]:

```
0.9666666666666667
```

In [26]:

```
accuracy_score(y_test, y_pred)
```

Out[26]:

```
0.9666666666666667
```

In [ ]:

```

```