In [1]:
```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

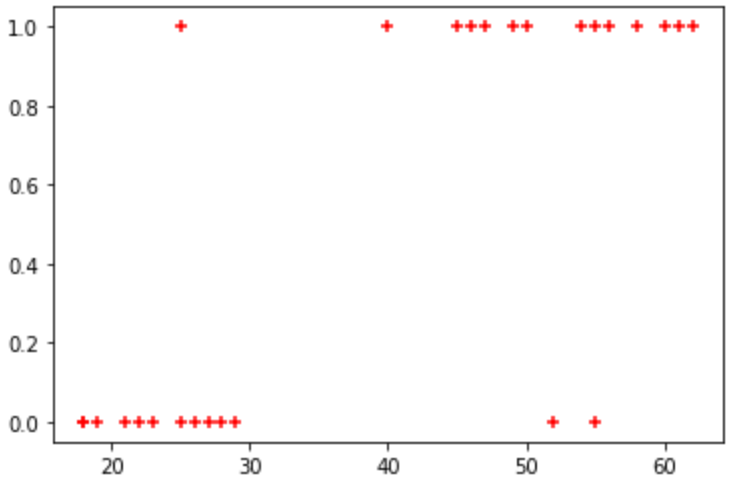
In [2]:
```
df = pd.read_csv(r"C:\Users\rishu\Desktop\data\insurance_data.csv")
df.head()
```

Out[2]:

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

In [3]:
```
plt.scatter(df.age,df.bought_insurance,marker='+',color='red')
```

Out[3]:
```
<matplotlib.collections.PathCollection at 0x1749f1b5f70>
```



In [6]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(df[['age']],df.bought_insurance,trai
```

In [7]:
```
X_test
```

Out[7]:

	age
17	58
14	49
18	19
3	52
23	45
10	18

In [8]:
```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

In [9]:
```
model.fit(X_train, y_train)
```

Out[9]:
```
LogisticRegression()
```

In [10]:
```
X_test
```

Out[10]:

	age
17	58
14	49
18	19
3	52
23	45
10	18

In [21]:
```
model.predict(X_test)
```

Out[21]:
```
array([1, 1, 0, 1, 1, 0], dtype=int64)
```

In [12]:
```
model.predict_proba(X_test)
```

Out[12]:
```
array([[0.04795632, 0.95204368],
       [0.15805894, 0.84194106],
       [0.93776133, 0.06223867],
       [0.10800643, 0.89199357],
       [0.25198743, 0.74801257],
       [0.94576637, 0.05423363]])
```

In [22]:
```
y_predicted=model.score(X_test,y_test)
```

In [14]:
```
y_predicted
```

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

In [15]:
```
model.coef_
```

Out[15]:
```
array([[0.14617537]])
```

In [16]:
```
model.intercept_
```

Out[16]:
```
array([-5.48985094])
```

Lets defined sigmoid function now and do the math with hand

In [17]:
```
import math
def sigmoid(x):
  return 1 / (1 + math.exp(-x))
```

In [18]:
```
def prediction_function(age):
    z = 0.042 * age - 1.53 # 0.04150133 ~ 0.042 and -1.52726963 ~ -1.53
    y = sigmoid(z)
    return y
```

In [19]:
```
age = 35
prediction_function(age)
```

Out[19]:
```
0.4850044983805899
```

0.485 is less than 0.5 which means person with 35 age will not buy insurance

In [20]:
```
age = 43
prediction_function(age)
```

Out[20]:
```
0.568565299077705
```

0.485 is more than 0.5 which means person with 43 will buy the insurance

Exercise Download employee retention dataset from here: https://www.kaggle.com/giripujar/hr-analytics.

Now do some exploratory data analysis to figure out which variables have direct and clear impact on employee retention (i.e. whether they leave the company or continue to work) Plot bar charts showing impact of employee salaries on retention Plot bar charts showing corelation between department and employee retention Now build logistic regression model using variables that were narrowed down in step 1 Measure the accuracy of the model

In [ ]:
```
```