In [1]:

```
import pandas as pd
```

In [3]:

```
df=pd.read_csv(r"D:\machine_learning\csv_files\gender_weight-height.csv")
```

In [4]:

```
df
```

Out[4]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...	...	...	...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

10000 rows × 3 columns

In [5]:

```
x=df.drop(["Gender","Weight"],axis=1)
y=df.Gender
```

In [6]:

```
from sklearn.model_selection import train_test_split
```

In [7]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y)
```

In [8]:

```
from sklearn.linear_model import LogisticRegression
```

In [9]:

```
model=LogisticRegression()
```

In [10]:

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

Out[10]:

```
LogisticRegression()
```

In [11]:

```
from sklearn.metrics import accuracy_score,confusion_matrix
```

In [12]:

```
pre=model.predict(x_test)
```

In [13]:

```
accuracy_score(pre,y_test)
```

Out[13]:

```
0.8348
```

In [14]:

```
confusion_matrix(pre,y_test)
```

Out[14]:

```
array([[1082,  207],
       [ 206, 1005]], dtype=int64)
```

In [15]:

```
import numpy as np
```

In [16]:

```
model.coef_
```

Out[16]:

```
array([[0.67965691]])
```

In [17]:

```
model.intercept_
```

Out[17]:

```
array([-45.07146433])
```

In [18]:

```
from scipy.special import expit
```

In [19]:

```
import matplotlib.pyplot as plt
```

In [31]:

```
model.predict([[65]])
```

Out[31]:

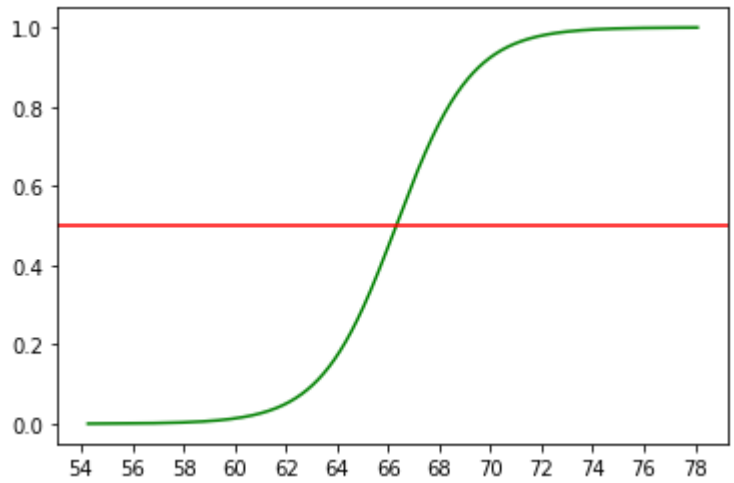
```
array(['Female'], dtype=object)
```

In [ ]:

```

```

In [30]:

```
z=np.sort(np.array(x_test),axis=0)
a = range(54, 79,2)

y_test =z* model.coef_ + model.intercept_
sigmoid = expit(np.array(y_test))
plt.plot(z,sigmoid,c="green", label = "logistic fit")
plt.axhline(.5, color="red", label="cutoff")
plt.xticks(ticks=a,label=a)

plt.show()
```



In [24]:

```
z
```

Out[24]:

```
array([[54.26313333],
       [55.73973682],
       [55.85121382],
       ...,
       [76.70983486],
       [77.44661995],
       [78.09586747]])
```

In [26]:

```
model.predict([[33]])
```

Out[26]:

```
array(['Female'], dtype=object)
```

In [53]:

```
M=df.drop(["Gender"],axis=1)
n=df.Gender
```

In [54]:

```
M
```

Out[54]:

	Height	Weight
0	73.847017	241.893563
1	68.781904	162.310473
2	74.110105	212.740856
3	71.730978	220.042470
4	69.881796	206.349801
...	...	...
9995	66.172652	136.777454
9996	67.067155	170.867906
9997	63.867992	128.475319
9998	69.034243	163.852461
9999	61.944246	113.649103

10000 rows × 2 columns

In [55]:

```
n
```

Out[55]:

```
0         Male
1         Male
2         Male
3         Male
4         Male
         ...
9995    Female
9996    Female
9997    Female
9998    Female
9999    Female
Name: Gender, Length: 10000, dtype: object
```

In [56]:

```
from sklearn.model_selection import train_test_split
M_train,M_test,n_train,n_test = train_test_split(M, n, random_state=0)
M_train
```

Out[56]:

	Height	Weight
2967	68.058837	187.779075
700	69.760095	187.812062
3481	71.702360	214.787698
1621	71.096113	210.821194
800	72.215035	204.937760
...	...	...
9225	60.421255	97.263881
4859	66.730755	174.156893
3264	67.467086	162.475957
9845	62.127480	136.783022
2732	70.597025	188.450674

7500 rows × 2 columns

In [57]:

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression()
classifier.fit(M_train,n_train)
```

Out[57]:

```
LogisticRegression()
```

In [58]:

```
z_pred = classifier.predict(M_test)
z_pred
```

Out[58]:

```
array(['Female', 'Male', 'Male', ..., 'Male', 'Male', 'Female'],
      dtype=object)
```

In [59]:

```
classifier.predict([[70.00, 170]])
```

Out[59]:

```
array(['Female'], dtype=object)
```

In [60]:

```
from sklearn.metrics import confusion_matrix
```

In [61]:

```
confusion_matrix(z_pred, n_test)
```

Out[61]:

```
array([[1140,  107],
       [  97, 1156]], dtype=int64)
```

In [62]:

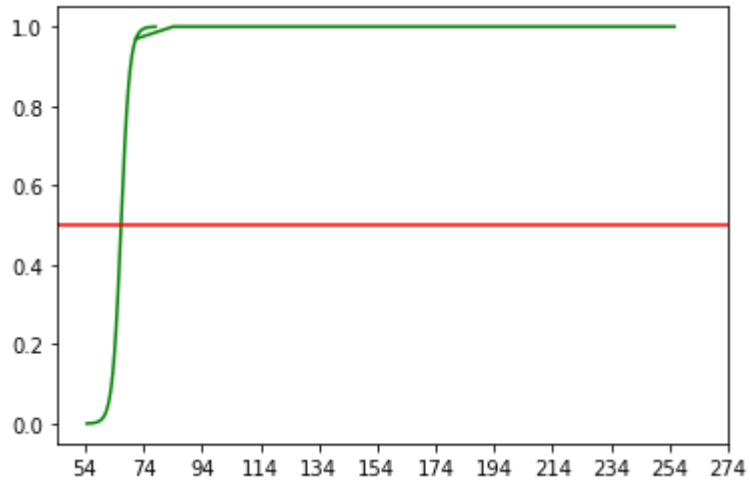
```
accuracy = (1140+1156)/(1140+1156+97+107)
accuracy
```

Out[62]:

```
0.9184
```

In [66]:

```
s=np.sort(np.array(M_test),axis=0)
a = range(54, 275,20)

p_test =s* model.coef_ + model.intercept_
sigmoid = expit(np.array(p_test))
plt.plot(s,sigmoid,c="green", label = "logistic fit")
plt.axhline(.5, color="red", label="cutoff")
plt.xticks(ticks=a,label=a)

plt.show()
```



In [67]:

```
s
```

Out[67]:

```
array([[ 54.61685783,  71.39374874],
       [ 55.73973682,  84.17069477],
       [ 56.63041198,  86.89071184],
       ...,
       [ 77.1008721 , 249.9462832 ],
       [ 77.4655691 , 252.55668944],
       [ 78.09586747, 255.69083484]])
```

In [69]:

```
p_test
```

Out[69]:

```
array([[ -7.95073976,   3.45178999],
       [ -7.1875673 ,  12.13572959],
       [ -6.58221378,  13.98440797],
       ...,
       [  7.33067579, 124.80625298],
       [  7.57854463, 126.58043361],
       [  8.00693127, 128.71057717]])
```

In [ ]:

```

```