

# 1) LOGISTIC REGRESSION MODEL ¶

In [27]:

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, accuracy_score
```

In [28]:

```
import warnings
warnings.filterwarnings('ignore')
```

## DATASET USED IS LUNG CANCER

In [29]:

```
df= pd.read_csv('/Users/anjali98/lc.csv')
df.head()
df.describe()
```

Out[29]:

	Age	Smokes	Alcohol	Result
count	107.000000	107.000000	107.000000	107.000000
mean	43.635514	16.046729	3.280374	0.485981
std	15.229931	6.724193	2.188281	0.502155
min	18.000000	0.000000	0.000000	0.000000
25%	32.000000	12.000000	2.000000	0.000000
50%	42.000000	20.000000	3.000000	0.000000
75%	56.000000	20.000000	4.500000	1.000000
max	77.000000	34.000000	8.000000	1.000000

In [30]:

```
print("Number of null values in the data set are - ",df.isnull().values.any().sum())
```

Number of null values in the data set are - 0

In [31]:

```
from sklearn.model_selection import train_test_split
y= df['Result']
x= df.drop(['Result'], axis=1)
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size= 0.4,random_state=42)
```

In [32]:

```
modelLogistic = LogisticRegression()  
modelLogistic.fit(x_train,y_train)
```

Out[32]:

```
LogisticRegression()
```

In [33]:

```
print("The intercept b0= ", modelLogistic.intercept_)  
print("The coefficient b1= ", modelLogistic.coef_)
```

```
The intercept b0= [-20.40018975]  
The coefficient b1= [[0.24838255 0.24177754 1.70960268]]
```

In [34]:

```
ypred = modelLogistic.predict(x_test)
```

In [35]:

```
log_score = modelLogistic.score(x_test, y_test)  
  
from sklearn import metrics  
cm = metrics.confusion_matrix(y_test, ypred)  
print(cm)
```

```
[[23  0]  
 [ 1 19]]
```

## FINAL ACCURACY SCORE

In [36]:

```
all_sample_title = 'Accuracy Score: {0}'.format(log_score)  
print(all_sample_title)
```

```
Accuracy Score: 0.9767441860465116
```

In [ ]: