

In [1]:

```
#importing necesary libraries
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.model_selection import cross_val_score
```

In [2]:

```
# Defining names of columns of data set
names = ['pelvic incidence', 'pelvic tilt', 'lumbar lordosis angle',
         'sacral slope', 'pelvic radius', 'grade of spondylolisthesis', 'label']

# read data
df = pd.read_csv('column_2C.dat',
                 sep=' ',
                 names=names,
                 header=None,
                 index_col = None,
                 na_values = '?')

#There are no missing values
#Shuffling data ser
df = df.sample(frac = 1)

#printing shape and first 10 rows of data set
print(df.shape)
df.head(10)
```

(310, 7)

Out[2]:

	pelvic incidence	pelvic tilt	lumbar lordosis angle	sacral slope	pelvic radius	grade of spondylolisthesis	label
22	63.07	24.41	54.00	38.66	106.42	15.78	AB
10	49.71	13.04	31.33	36.67	108.65	-7.83	AB
171	78.40	14.04	79.69	64.36	104.73	12.39	AB
156	79.48	26.73	70.65	52.74	118.59	61.70	AB
0	63.03	22.55	39.61	40.48	98.67	-0.25	AB
85	45.44	9.91	45.00	35.54	163.07	20.32	AB
90	81.66	28.75	58.23	52.91	114.77	30.61	AB
225	59.73	7.72	55.34	52.00	125.17	3.24	NO
17	31.28	3.14	32.56	28.13	129.01	3.62	AB
114	80.99	36.84	86.96	44.14	141.09	85.87	AB

In [3]:

```
#Describing basic features of dataset
df.describe()
```

Out[3]:

	pelvic incidence	pelvic tilt	lumbar lordosis angle	sacral slope	pelvic radius	grade of spondylolisthesis
count	310.000000	310.000000	310.000000	310.000000	310.000000	310.000000
mean	60.496484	17.542903	51.930710	42.953871	117.920548	26.296742
std	17.236109	10.008140	18.553766	13.422748	13.317629	37.558883
min	26.150000	-6.550000	14.000000	13.370000	70.080000	-11.060000
25%	46.432500	10.667500	37.000000	33.347500	110.710000	1.600000

50%	58.690000 pelvic incidence	16.360000 pelvic tilt	lumbosacral angle	42.450000 sacral slope	118.265000 pelvic radius	1 grade of spondylolisthesis
75%	72.880000	22.120000	63.000000	52.692500	125.467500	41.285000
max	129.830000	49.430000	125.740000	121.430000	163.070000	418.540000

In [4]:

```
#dividing into input and labels
X = df.iloc[:, :6]
Y = df.iloc[:, 6]
```

In [5]:

```
# build and fit model
reg = LogisticRegression(solver = 'lbfgs')
reg.fit(X,Y)
```

Out[5]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=100,
                    multi_class='warn', n_jobs=None, penalty='l2',
                    random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                    warm_start=False)
```

In [6]:

```
# cross-validate
k = 10
scores = cross_val_score(estimator=reg,
                          X=X,
                          y=Y,
                          scoring="accuracy",
                          cv=k,
                          )

acc = scores.mean()
print("Accuracy calculated using %d-fold cross validation = %.3f" % (k, acc))
```

Accuracy calculated using 10-fold cross validation = 0.855

In [9]:

```
print('The program will show you prediction if the patient has abnormalities in vertebral column o
r no.\n',
      'The accuracy of prediction is %.3f ' % acc)
cont = 'yes'
while (cont == 'yes'):
    I = np.array([[0,0,0,0,0,0]])
    print('\nPlease enter the values of following parameters:\n')
    I[0,0] = float(input(names[0] + ' '))
    I[0,1] = float(input(names[1] + ' '))
    I[0,2] = float(input(names[2] + ' '))
    I[0,3] = float(input(names[3] + ' '))
    I[0,4] = float(input(names[4] + ' '))
    I[0,5] = float(input(names[5] + ' '))

    ypred = reg.predict(I)
    if ypred == 'AB':
        print( '*****',
              '\nPatient has abnormalities in vertebral column\n',
              '\n*****')
    else:
        print('\n*****',
              '\nPatient is healthy\n',
              '\n*****')
    cont = input('Do you want to ask about next patient?[yes/no] ')
```

The program will show you prediction if the patient has abnormalities in vertebral column or no.
The accuracy of prediction is 0.855

Plase enter the values of folowing parameters:

pelvic incidence 59.73
pelvic tilt 7.7
lumbar lordosis angle 55
sacral slope 52
pelvic radius 125
grade of spondylolisthesis 3.34

Patient is healthy

Do you want to ask about next patient?[yes/no] yes

Plase enter the values of folowing parameters:

pelvic incidence 80.9
pelvic tilt 36
lumbar lordosis angle 86.9
sacral slope 44
pelvic radius 141
grade of spondylolisthesis 85.8

Patient has abnormalities in vertebral column

Do you want to ask about next patient?[yes/no] no

In []:

In []: