Logistic Regression -1

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
from sklearn.linear_model import LogisticRegression
```

In [3]:
```
df=pd.read_csv("1_ionosphere.csv")
df
```

Out[3]:

	1	0	0.99539	-0.05889	0.85243	0.02306	0.83398	-0.37708	1.1	0.03760	...	-0.51171	C
0	1	0	1.00000	-0.18829	0.93035	-0.36156	-0.10868	-0.93597	1.00000	-0.04549	...	-0.26569	-(
1	1	0	1.00000	-0.03365	1.00000	0.00485	1.00000	-0.12062	0.88965	0.01198	...	-0.40220	(
2	1	0	1.00000	-0.45161	1.00000	1.00000	0.71216	-1.00000	0.00000	0.00000	...	0.90695	(
3	1	0	1.00000	-0.02401	0.94140	0.06531	0.92106	-0.23255	0.77152	-0.16399	...	-0.65158	(
4	1	0	0.02337	-0.00592	-0.09924	-0.11949	-0.00763	-0.11824	0.14706	0.06637	...	-0.01535	-(
...	...	...	...	...	...	...	...	...	...	...	...	...	
345	1	0	0.83508	0.08298	0.73739	-0.14706	0.84349	-0.05567	0.90441	-0.04622	...	-0.04202	(
346	1	0	0.95113	0.00419	0.95183	-0.02723	0.93438	-0.01920	0.94590	0.01606	...	0.01361	(
347	1	0	0.94701	-0.00034	0.93207	-0.03227	0.95177	-0.03431	0.95584	0.02446	...	0.03193	(
348	1	0	0.90608	-0.01657	0.98122	-0.01989	0.95691	-0.03646	0.85746	0.00110	...	-0.02099	(
349	1	0	0.84710	0.13533	0.73638	-0.06151	0.87873	0.08260	0.88928	-0.09139	...	-0.15114	(

350 rows × 35 columns

In [4]:
```
feature_matrix=df.iloc[:,0:34]
target_vector=df.iloc[:,-1]
```

In [5]:
```
feature_matrix.shape
```

Out[5]: (350, 34)

In [6]:
```
target_vector.shape
```

Out[6]: (350,)

In [7]:
```
from sklearn.preprocessing import StandardScaler
```

In [10]:
```
fs=StandardScaler().fit_transform(feature_matrix)
```

In [11]:
```
logr=LogisticRegression()
logr.fit(fs,target_vector)
```

Out[11]:
```
LogisticRegression()
```

In [12]:
```
observation=[[1.4,2.3,-5.9,11,12,13,14,15,16,17,1,2,3,4,5,6,7,8,9,10,21,22,23,24,25,26,
```

In [15]:
```
prediction=logr.predict(observation)
print(prediction)
```

```
['g']
```

In [16]:
```
logr.classes_
```

Out[16]:
```
array(['b', 'g'], dtype=object)
```

In [18]:
```
logr.predict_proba(observation)[0][0]
```

Out[18]:
```
0.0
```

In [19]:
```
logr.predict_proba(observation)[0][1]
```

Out[19]:
```
1.0
```

Logistic Regression - 2

In [20]:
```
import re
from sklearn.datasets import load_digits
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
```

In [21]:
```
digits=load_digits()
digits
```

Out[21]:
```
{'data': array([[ 0.,  0.,  5., ...,  0.,  0.,  0.],
       [ 0.,  0.,  0., ..., 10.,  0.,  0.],
       [ 0.,  0.,  0., ..., 16.,  9.,  0.],
```

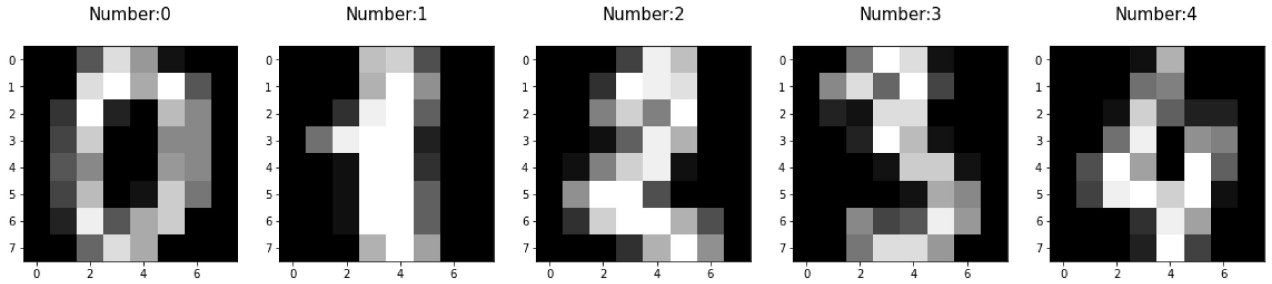
```
       ...,
       [ 0.,  0.,  1., ...,  6.,  0.,  0.],
       [ 0.,  0.,  2., ..., 12.,  0.,  0.],
       [ 0.,  0., 10., ..., 12.,  1.,  0.]]),
'target': array([0, 1, 2, ..., 8, 9, 8]),
'frame': None,
'feature_names': ['pixel_0_0',
 'pixel_0_1',
 'pixel_0_2',
 'pixel_0_3',
 'pixel_0_4',
 'pixel_0_5',
 'pixel_0_6',
 'pixel_0_7',
 'pixel_1_0',
 'pixel_1_1',
 'pixel_1_2',
 'pixel_1_3',
 'pixel_1_4',
 'pixel_1_5',
 'pixel_1_6',
 'pixel_1_7',
 'pixel_2_0',
 'pixel_2_1',
 'pixel_2_2',
 'pixel_2_3',
 'pixel_2_4',
 'pixel_2_5',
 'pixel_2_6',
 'pixel_2_7',
 'pixel_3_0',
 'pixel_3_1',
 'pixel_3_2',
 'pixel_3_3',
 'pixel_3_4',
 'pixel_3_5',
 'pixel_3_6',
 'pixel_3_7',
 'pixel_4_0',
 'pixel_4_1',
 'pixel_4_2',
 'pixel_4_3',
 'pixel_4_4',
 'pixel_4_5',
 'pixel_4_6',
 'pixel_4_7',
 'pixel_5_0',
 'pixel_5_1',
 'pixel_5_2',
 'pixel_5_3',
 'pixel_5_4',
 'pixel_5_5',
 'pixel_5_6',
 'pixel_5_7',
 'pixel_6_0',
 'pixel_6_1',
 'pixel_6_2',
 'pixel_6_3',
 'pixel_6_4',
 'pixel_6_5',
 'pixel_6_6',
 'pixel_6_7',
 'pixel_7_0',
 'pixel_7_1',
 'pixel_7_2',
```

```
 'pixel_7_3',
 'pixel_7_4',
 'pixel_7_5',
 'pixel_7_6',
 'pixel_7_7'],
'target_names': array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
'images': array([[[ 0.,  0.,  5., ...,  1.,  0.,  0.],
        [ 0.,  0., 13., ..., 15.,  5.,  0.],
        [ 0.,  3., 15., ..., 11.,  8.,  0.],
        ...,
        [ 0.,  4., 11., ..., 12.,  7.,  0.],
        [ 0.,  2., 14., ..., 12.,  0.,  0.],
        [ 0.,  0.,  6., ...,  0.,  0.,  0.]],

       [[ 0.,  0.,  0., ...,  5.,  0.,  0.],
        [ 0.,  0.,  0., ...,  9.,  0.,  0.],
        [ 0.,  0.,  3., ...,  6.,  0.,  0.],
        ...,
        [ 0.,  0.,  1., ...,  6.,  0.,  0.],
        [ 0.,  0.,  1., ...,  6.,  0.,  0.],
        [ 0.,  0.,  0., ..., 10.,  0.,  0.]],

       [[ 0.,  0.,  0., ..., 12.,  0.,  0.],
        [ 0.,  0.,  3., ..., 14.,  0.,  0.],
        [ 0.,  0.,  8., ..., 16.,  0.,  0.],
        ...,
        [ 0.,  9., 16., ...,  0.,  0.,  0.],
        [ 0.,  3., 13., ..., 11.,  5.,  0.],
        [ 0.,  0.,  0., ..., 16.,  9.,  0.]],

       ...,

       [[ 0.,  0.,  1., ...,  1.,  0.,  0.],
        [ 0.,  0., 13., ...,  2.,  1.,  0.],
        [ 0.,  0., 16., ..., 16.,  5.,  0.],
        ...,
        [ 0.,  0., 16., ..., 15.,  0.,  0.],
        [ 0.,  0., 15., ..., 16.,  0.,  0.],
        [ 0.,  0.,  2., ...,  6.,  0.,  0.]],

       [[ 0.,  0.,  2., ...,  0.,  0.,  0.],
        [ 0.,  0., 14., ..., 15.,  1.,  0.],
        [ 0.,  4., 16., ..., 16.,  7.,  0.],
        ...,
        [ 0.,  0.,  0., ..., 16.,  2.,  0.],
        [ 0.,  0.,  4., ..., 16.,  2.,  0.],
        [ 0.,  0.,  5., ..., 12.,  0.,  0.]],

       [[ 0.,  0., 10., ...,  1.,  0.,  0.],
        [ 0.,  2., 16., ...,  1.,  0.,  0.],
        [ 0.,  0., 15., ..., 15.,  0.,  0.],
        ...,
        [ 0.,  4., 16., ..., 16.,  6.,  0.],
        [ 0.,  8., 16., ..., 16.,  8.,  0.],
        [ 0.,  1.,  8., ..., 12.,  1.,  0.]]]),
 'DESCR': ".. _digits_dataset:\n\nOptical recognition of handwritten digits dataset\n---
----------------------------------------------\n\n**Data Set Characteristics:**\n\n
:Number of Instances: 1797\n    :Number of Attributes: 64\n    :Attribute Information: 8
x8 image of integer pixels in the range 0..16.\n    :Missing Attribute Values: None\n
:Creator: E. Alpaydin (alpaydin '@' boun.edu.tr)\n    :Date: July; 1998\n\nThis is a cop
y of the test set of the UCI ML hand-written digits datasets\nhttps://archive.ics.uci.ed
u/ml/datasets/Optical+Recognition+of+Handwritten+Digits\n\nThe data set contains images
of hand-written digits: 10 classes where\neach class refers to a digit.\n\nPreprocessing
programs made available by NIST were used to extract\nnormalized bitmaps of handwritten
digits from a preprinted form. From a\ntotal of 43 people, 30 contributed to the trainin
```

```
g set and different 13\nto the test set. 32x32 bitmaps are divided into nonoverlapping b
locks of\n4x4 and the number of on pixels are counted in each block. This generates\nan
input matrix of 8x8 where each element is an integer in the range\n0..16. This reduces d
imensionality and gives invariance to small\ndistortions.\n\nFor info on NIST preprocess
ing routines, see M. D. Garris, J. L. Blue, G.\nT. Candela, D. L. Dimmick, J. Geist, P.
J. Grother, S. A. Janet, and C.\nL. Wilson, NIST Form-Based Handprint Recognition Syste
m, NISTIR 5469,\n1994.\n\n.. topic:: References\n\n  - C. Kaynak (1995) Methods of Combi
ning Multiple Classifiers and Their\n    Applications to Handwritten Digit Recognition,
MSc Thesis, Institute of\n    Graduate Studies in Science and Engineering, Bogazici Univ
ersity.\n  - E. Alpaydin, C. Kaynak (1998) Cascading Classifiers, Kybernetika.\n  - Ken
Tang and Ponnuthurai N. Suganthan and Xi Yao and A. Kai Qin.\n    Linear dimensionalityr
eduction using relevance weighted LDA. School of\n    Electrical and Electronic Engineer
ing Nanyang Technological University.\n    2005.\n  - Claudio Gentile. A New Approximate
Maximal Margin Classification\n    Algorithm. NIPS. 2000.\n"}
```

In [26]:

```
plt.figure(figsize=(20,4))
for index,(image,label) in enumerate(zip(digits.data[0:5],digits.target[0:5])):
    plt.subplot(1,5,index+1)
    plt.imshow(np.reshape(image,(8,8)),cmap=plt.cm.gray)
    plt.title('Number:%i\n'%label,fontsize=15)
```



In [27]:

```
x_train,x_test,y_train,y_test=train_test_split(digits.data,digits.target,test_size=0.30
```

In [28]:

```
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(1257, 64)
(540, 64)
(1257,)
(540,)
```

In [30]:

```
logre=LogisticRegression(max_iter=10000)
logre.fit(x_train,y_train)
```

Out[30]:

```
LogisticRegression(max_iter=10000)
```

In [31]:

```
print(logre.predict(x_test))
```

```
[4 2 5 5 4 3 6 8 2 6 2 4 7 7 2 4 5 0 1 1 5 9 2 3 3 1 2 0 0 5 6 0 3 4 3 1 7
 8 5 5 0 4 6 5 4 5 3 1 2 3 4 3 2 9 7 8 5 5 2 9 7 7 5 4 2 6 1 2 1 8 2 3 6 0
 6 3 7 9 9 7 3 8 8 0 8 8 1 9 7 1 8 9 2 8 7 6 5 5 4 6 3 1 5 1 5 6 7 7 7 1 0
 0 8 7 4 1 3 1 1 9 1 7 5 3 7 6 5 7 7 3 9 3 7 2 1 1 7 4 3 0 5 8 7 8 1 8 3 2
 3 6 6 6 4 4 8 2 0 6 6 3 3 8 9 9 6 1 2 9 6 9 2 7 9 1 0 3 0 7 9 0 2 2 4 9 6
 1 0 4 5 1 5 6 0 4 8 3 7 0 4 1 7 2 7 9 4 5 6 5 2 5 4 8 1 2 5 8 1 5 3 6 4 6
 6 8 6 3 5 0 4 7 2 0 4 1 8 8 8 2 9 9 1 7 5 9 4 3 3 1 3 0 6 7 5 2 2 6 2 6 8
```

```
6 1 1 8 9 0 8 2 3 1 7 8 6 2 7 5 2 7 2 9 1 3 5 3 2 9 8 8 6 6 1 5 6 6 9 8 6
7 8 4 6 1 6 3 5 7 6 3 0 2 4 1 8 8 7 1 4 2 7 7 5 0 7 5 4 8 1 3 9 2 4 8 9 1
4 7 3 2 9 0 9 2 4 4 2 5 8 3 6 6 7 8 2 0 8 4 5 7 5 2 7 4 4 7 7 5 9 2 8 6 0
5 7 8 4 9 3 8 3 6 4 2 4 7 4 6 9 2 1 9 9 0 1 6 2 4 0 7 8 4 6 2 9 2 6 0 6 4
9 5 0 3 0 3 7 3 2 4 8 9 2 3 7 4 5 6 3 8 3 1 6 8 1 0 8 6 3 1 9 0 4 7 9 6 3
2 5 4 1 3 8 2 8 5 1 3 4 0 7 9 0 3 4 4 9 4 3 7 5 0 4 9 2 3 9 7 8 1 6 9 1 0
8 4 6 2 9 0 6 1 4 2 2 9 6 5 8 0 0 7 0 1 1 1 0 2 2 4 2 3 4 0 4 6 0 3 9 0 3
3 4 1 5 5 4 2 3 3 7 1 3 6 4 7 6 0 9 2 7 9 0]
```

In [32]:
```
print(logre.score(x_test,y_test))
```

```
0.9592592592592593
```

Random Forest

In [4]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:
```
df=pd.read_csv("1_ionosphere.csv")
```

In [8]:
```
df['g'].value_counts()
```

Out[8]:
```
g    224
b    126
Name: g, dtype: int64
```

In [9]:
```
x=df.drop('g',axis=1)
y=df['g']
```

In [10]:
```
g1={"g":{"g":1,"b":2}}
df=df.replace(g1)
df
```

Out[10]:

	1	0	0.99539	-0.05889	0.85243	0.02306	0.83398	-0.37708	1.1	0.03760	...	-0.51171	C
0	1	0	1.00000	-0.18829	0.93035	-0.36156	-0.10868	-0.93597	1.00000	-0.04549	...	-0.26569	-(
1	1	0	1.00000	-0.03365	1.00000	0.00485	1.00000	-0.12062	0.88965	0.01198	...	-0.40220	(
2	1	0	1.00000	-0.45161	1.00000	1.00000	0.71216	-1.00000	0.00000	0.00000	...	0.90695	(
3	1	0	1.00000	-0.02401	0.94140	0.06531	0.92106	-0.23255	0.77152	-0.16399	...	-0.65158	(
4	1	0	0.02337	-0.00592	-0.09924	-0.11949	-0.00763	-0.11824	0.14706	0.06637	...	-0.01535	-(
...	...	...	...	...	...	...	...	...	...	...	...	...	
345	1	0	0.83508	0.08298	0.73739	-0.14706	0.84349	-0.05567	0.90441	-0.04622	...	-0.04202	(
346	1	0	0.95113	0.00419	0.95183	-0.02723	0.93438	-0.01920	0.94590	0.01606	...	0.01361	(
347	1	0	0.94701	-0.00034	0.93207	-0.03227	0.95177	-0.03431	0.95584	0.02446	...	0.03193	(

	1	0	0.99539	-0.05889	0.85243	0.02306	0.83398	-0.37708	1.1	0.03760	...	-0.51171	C
348	1	0	0.90608	-0.01657	0.98122	-0.01989	0.95691	-0.03646	0.85746	0.00110	...	-0.02099	
349	1	0	0.84710	0.13533	0.73638	-0.06151	0.87873	0.08260	0.88928	-0.09139	...	-0.15114	

350 rows × 35 columns

In [11]:
```
from sklearn.model_selection import train_test_split
```

In [12]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

In [13]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [14]:
```
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[14]:
```
RandomForestClassifier()
```

In [24]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]
}
```

In [25]:
```
from sklearn.model_selection import GridSearchCV
grid_search =GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

Out[25]:
```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 10, 15, 20, 25],
                         'n_estimators': [10, 20, 30, 40, 50]},
             scoring='accuracy')
```

In [26]:
```
grid_search.best_score_
```

Out[26]:
```
0.9385245901639344
```

In [28]:
```
rfc_best=grid_search.best_estimator_
```

In [30]:
```
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','No'],fill
```

Out[30]:
```
[Text(1488.0, 1956.96, '0.85243 <= 0.033\ngini = 0.461\nsamples = 159\nvalue = [88, 156]
\nclass = No'),
```

```
 Text(992.0, 1522.0800000000002, 'gini = 0.0\nsamples = 33\nvalue = [54, 0]\nclass = Ye
s'),
 Text(1984.0, 1522.0800000000002, '-0.51171 <= -0.981\ngini = 0.294\nsamples = 126\nvalu
e = [34, 156]\nclass = No'),
 Text(992.0, 1087.2, '0.59755 <= 0.5\ngini = 0.219\nsamples = 13\nvalue = [14, 2]\nclass
= Yes'),
 Text(496.0, 652.3200000000002, 'gini = 0.48\nsamples = 5\nvalue = [3, 2]\nclass = Ye
s'),
 Text(1488.0, 652.3200000000002, 'gini = 0.0\nsamples = 8\nvalue = [11, 0]\nclass = Ye
s'),
 Text(2976.0, 1087.2, '-0.17755 <= -0.749\ngini = 0.203\nsamples = 113\nvalue = [20, 15
4]\nclass = No'),
 Text(2480.0, 652.3200000000002, 'gini = 0.278\nsamples = 5\nvalue = [5, 1]\nclass = Ye
s'),
 Text(3472.0, 652.3200000000002, '0.59755 <= 0.999\ngini = 0.163\nsamples = 108\nvalue =
[15, 153]\nclass = No'),
 Text(2976.0, 217.44000000000005, 'gini = 0.091\nsamples = 93\nvalue = [7, 139]\nclass =
No'),
 Text(3968.0, 217.44000000000005, 'gini = 0.463\nsamples = 15\nvalue = [8, 14]\nclass =
No')]
```