# Class_16.04.2023_Naive_Bayes_Classifier

April 16, 2023

We will use the breast cancer here to look into naive bayes classifier application

[105]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.datasets import load_breast_cancer
```

[106]:
```
cancer = load_breast_cancer()
```

[107]:
```
print(f'the attributes in the data are {cancer.feature_names}')
print(f'the target section in the data are {cancer.target_names}')
```

```
the attributes in the data are ['mean radius' 'mean texture' 'mean perimeter'
'mean area'
 'mean smoothness' 'mean compactness' 'mean concavity'
 'mean concave points' 'mean symmetry' 'mean fractal dimension'
 'radius error' 'texture error' 'perimeter error' 'area error'
 'smoothness error' 'compactness error' 'concavity error'
 'concave points error' 'symmetry error' 'fractal dimension error'
 'worst radius' 'worst texture' 'worst perimeter' 'worst area'
 'worst smoothness' 'worst compactness' 'worst concavity'
 'worst concave points' 'worst symmetry' 'worst fractal dimension']
the target section in the data are ['malignant' 'benign']
```

[108]:
```
# converting it into a pandas dataframe
cancer_df = pd.DataFrame(data = cancer.data , columns=(cancer.feature_names))
```

[109]:
```
cancer_df.tail()
```

[109]:

| | mean radius | mean texture | mean perimeter | mean area | mean smoothness | \ |
|---|---|---|---|---|---|---|
| 564 | 21.56 | 22.39 | 142.00 | 1479.0 | 0.11100 | |
| 565 | 20.13 | 28.25 | 131.20 | 1261.0 | 0.09780 | |
| 566 | 16.60 | 28.08 | 108.30 | 858.1 | 0.08455 | |
| 567 | 20.60 | 29.33 | 140.10 | 1265.0 | 0.11780 | |
| 568 | 7.76 | 24.54 | 47.92 | 181.0 | 0.05263 | |

| | mean compactness | mean concavity | mean concave points | mean symmetry | \ |
|---|---|---|---|---|---|
| 564 | 0.11590 | 0.24390 | 0.13890 | 0.1726 | |
| 565 | 0.10340 | 0.14400 | 0.09791 | 0.1752 | |

```
566          0.10230         0.09251              0.05302         0.1590
567          0.27700         0.35140              0.15200         0.2397
568          0.04362         0.00000              0.00000         0.1587

     mean fractal dimension  …  worst radius  worst texture  \
564                 0.05623  …        25.450          26.40
565                 0.05533  …        23.690          38.25
566                 0.05648  …        18.980          34.12
567                 0.07016  …        25.740          39.42
568                 0.05884  …         9.456          30.37

     worst perimeter  worst area  worst smoothness  worst compactness  \
564           166.10      2027.0           0.14100            0.21130
565           155.00      1731.0           0.11660            0.19220
566           126.70      1124.0           0.11390            0.30940
567           184.60      1821.0           0.16500            0.86810
568            59.16       268.6           0.08996            0.06444

     worst concavity  worst concave points  worst symmetry  \
564           0.4107                0.2216          0.2060
565           0.3215                0.1628          0.2572
566           0.3403                0.1418          0.2218
567           0.9387                0.2650          0.4087
568           0.0000                0.0000          0.2871

     worst fractal dimension
564                  0.07115
565                  0.06637
566                  0.07820
567                  0.12400
568                  0.07039

[5 rows x 30 columns]
```

[110]:
```
cancer_df['target'] = cancer.target
```

[111]:
```
cancer_df.head()
```

[111]:
```
   mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
0        17.99         10.38          122.80     1001.0          0.11840
1        20.57         17.77          132.90     1326.0          0.08474
2        19.69         21.25          130.00     1203.0          0.10960
3        11.42         20.38           77.58      386.1          0.14250
4        20.29         14.34          135.10     1297.0          0.10030

   mean compactness  mean concavity  mean concave points  mean symmetry  \
0           0.27760          0.3001              0.14710         0.2419
```

```
1           0.07864          0.0869              0.07017         0.1812
2           0.15990          0.1974              0.12790         0.2069
3           0.28390          0.2414              0.10520         0.2597
4           0.13280          0.1980              0.10430         0.1809

   mean fractal dimension  …  worst texture  worst perimeter  worst area  \
0                 0.07871  …          17.33           184.60      2019.0
1                 0.05667  …          23.41           158.80      1956.0
2                 0.05999  …          25.53           152.50      1709.0
3                 0.09744  …          26.50            98.87       567.7
4                 0.05883  …          16.67           152.20      1575.0

   worst smoothness  worst compactness  worst concavity  worst concave points  \
0            0.1622             0.6656           0.7119                0.2654
1            0.1238             0.1866           0.2416                0.1860
2            0.1444             0.4245           0.4504                0.2430
3            0.2098             0.8663           0.6869                0.2575
4            0.1374             0.2050           0.4000                0.1625

   worst symmetry  worst fractal dimension  target
0          0.4601                  0.11890       0
1          0.2750                  0.08902       0
2          0.3613                  0.08758       0
3          0.6638                  0.17300       0
4          0.2364                  0.07678       0

[5 rows x 31 columns]
```

[112]:
```
x = cancer_df.drop('target',axis = 'columns')
y = cancer_df.target
```

[113]:
```
x
```

[113]:
```
     mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
0          17.99         10.38          122.80     1001.0          0.11840
1          20.57         17.77          132.90     1326.0          0.08474
2          19.69         21.25          130.00     1203.0          0.10960
3          11.42         20.38           77.58      386.1          0.14250
4          20.29         14.34          135.10     1297.0          0.10030
..           …             …               …          …               …
564        21.56         22.39          142.00     1479.0          0.11100
565        20.13         28.25          131.20     1261.0          0.09780
566        16.60         28.08          108.30      858.1          0.08455
567        20.60         29.33          140.10     1265.0          0.11780
568         7.76         24.54           47.92      181.0          0.05263

     mean compactness  mean concavity  mean concave points  mean symmetry  \
```

```
0             0.27760         0.30010              0.14710         0.2419
1             0.07864         0.08690              0.07017         0.1812
2             0.15990         0.19740              0.12790         0.2069
3             0.28390         0.24140              0.10520         0.2597
4             0.13280         0.19800              0.10430         0.1809
..                …               …                    …             …
564           0.11590         0.24390              0.13890         0.1726
565           0.10340         0.14400              0.09791         0.1752
566           0.10230         0.09251              0.05302         0.1590
567           0.27700         0.35140              0.15200         0.2397
568           0.04362         0.00000              0.00000         0.1587

     mean fractal dimension  …  worst radius  worst texture  \
0                   0.07871  …        25.380          17.33
1                   0.05667  …        24.990          23.41
2                   0.05999  …        23.570          25.53
3                   0.09744  …        14.910          26.50
4                   0.05883  …        22.540          16.67
..                      …  …           …             …
564                 0.05623  …        25.450          26.40
565                 0.05533  …        23.690          38.25
566                 0.05648  …        18.980          34.12
567                 0.07016  …        25.740          39.42
568                 0.05884  …         9.456          30.37

     worst perimeter  worst area  worst smoothness  worst compactness  \
0             184.60      2019.0           0.16220            0.66560
1             158.80      1956.0           0.12380            0.18660
2             152.50      1709.0           0.14440            0.42450
3              98.87       567.7           0.20980            0.86630
4             152.20      1575.0           0.13740            0.20500
..               …          …                …                  …
564           166.10      2027.0           0.14100            0.21130
565           155.00      1731.0           0.11660            0.19220
566           126.70      1124.0           0.11390            0.30940
567           184.60      1821.0           0.16500            0.86810
568            59.16       268.6           0.08996            0.06444

     worst concavity  worst concave points  worst symmetry  \
0             0.7119                0.2654          0.4601
1             0.2416                0.1860          0.2750
2             0.4504                0.2430          0.3613
3             0.6869                0.2575          0.6638
4             0.4000                0.1625          0.2364
..               …                     …               …
564           0.4107                0.2216          0.2060
565           0.3215                0.1628          0.2572
```

```
566          0.3403                0.1418          0.2218
567          0.9387                0.2650          0.4087
568          0.0000                0.0000          0.2871

     worst fractal dimension
0                    0.11890
1                    0.08902
2                    0.08758
3                    0.17300
4                    0.07678
..                       …
564                  0.07115
565                  0.06637
566                  0.07820
567                  0.12400
568                  0.07039

[569 rows x 30 columns]
```

[114]:
```
y
```

[114]:
```
0      0
1      0
2      0
3      0
4      0
      ..
564    0
565    0
566    0
567    0
568    1
Name: target, Length: 569, dtype: int32
```

[115]:
```
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.2)
```

[116]:
```
xtrain.shape
```

[116]:
```
(455, 30)
```

[117]:
```
xtest.shape
```

[117]:
```
(114, 30)
```

[118]:
```
ytrain.shape
```

[118]:
```
(455,)
```

[119]:
```
ytest.shape
```

[119]:
```
(114,)
```

[120]:
```
#model implementation
from sklearn.naive_bayes import GaussianNB
```

[121]:
```
nb = GaussianNB()
nb.fit(xtrain,ytrain)
```

[121]:
```
GaussianNB()
```

[122]:
```
nb.score(xtest,ytest)
```

[122]:
```
0.956140350877193
```

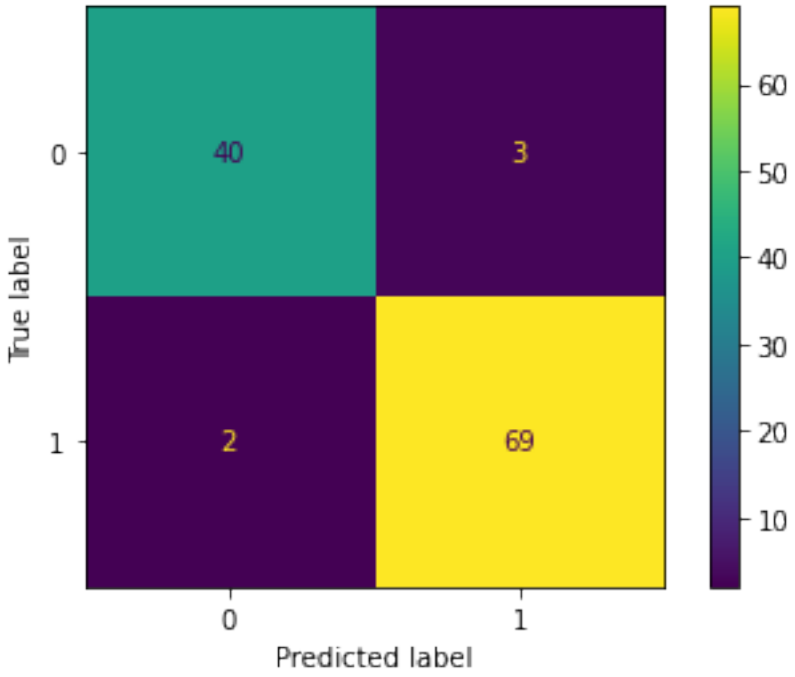
[123]:
```
from sklearn.model_selection import cross_val_score
from sklearn.metrics import accuracy_score, confusion_matrix ,␣
↪ConfusionMatrixDisplay
from sklearn.model_selection import GridSearchCV
```

[124]:
```
# Accuracy Test
''' Accuracy on training data'''
predict_train = nb.predict(xtrain)
accuracy_train = accuracy_score(ytrain,predict_train)
''' Accuracy on testing data'''
predict_test = nb.predict(xtest)
accuracy_test = accuracy_score(ytest,predict_test)
```

[125]:
```
print(f'accuracy on training dataset is: ',accuracy_train)
print(f'accuracy on testing dataset is: ',accuracy_test)
```

```
accuracy on training dataset is:  0.9362637362637363
accuracy on testing dataset is:  0.956140350877193
```

[126]:
```
# confusion matrix
import matplotlib.pyplot as plt
predictions = nb.predict(xtest)
cm = confusion_matrix(ytest, predictions, labels=clf.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm,display_labels=clf.classes_)
disp.plot()
plt.show()
```




[127]:
```
# finding the cross validation score
cv_score = cross_val_score(nb,x,y,cv = 3)
```

[128]:
```
print(cv_score)
```

```
[0.91578947 0.94736842 0.94708995]
```

[ ]:
```

```