# 4.1.SVM

May 15, 2023

[212]:

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import time
from subprocess import check_output
from scipy import stats
plt.style.use("ggplot")
import warnings
warnings.filterwarnings("ignore")
```

[213]:

```python
data=pd.read_csv('wdbc.data',header=None)
```

data.head()

[214]:

```python
data.head()
```

[214]:

```
          0  1      2      3       4       5        6        7       8  \
0    842302  M  17.99  10.38  122.80  1001.0  0.11840  0.27760  0.3001
1    842517  M  20.57  17.77  132.90  1326.0  0.08474  0.07864  0.0869
2  84300903  M  19.69  21.25  130.00  1203.0  0.10960  0.15990  0.1974
3  84348301  M  11.42  20.38   77.58   386.1  0.14250  0.28390  0.2414
4  84358402  M  20.29  14.34  135.10  1297.0  0.10030  0.13280  0.1980

         9  …     22     23      24      25      26      27      28      29  \
0  0.14710  …  25.38  17.33  184.60  2019.0  0.1622  0.6656  0.7119  0.2654
1  0.07017  …  24.99  23.41  158.80  1956.0  0.1238  0.1866  0.2416  0.1860
2  0.12790  …  23.57  25.53  152.50  1709.0  0.1444  0.4245  0.4504  0.2430
3  0.10520  …  14.91  26.50   98.87   567.7  0.2098  0.8663  0.6869  0.2575
4  0.10430  …  22.54  16.67  152.20  1575.0  0.1374  0.2050  0.4000  0.1625

       30       31
0  0.4601  0.11890
1  0.2750  0.08902
2  0.3613  0.08758
3  0.6638  0.17300
4  0.2364  0.07678
```

```
[5 rows x 32 columns]
```

[215]:

```
headers=['id','diagnosis','mean_radius','mean_texture','mean_perimeter','mean_area','mean_smoo
 ↪points','mean_symmetry','mean_fractal␣
 ↪dimension','SE_radius','SE_texture','SE_perimeter','SE_area','SE_smoothness','SE_compactnes
 ↪points','SE_symmetry','SE_fractal␣
 ↪dimension','worst_radius','worst_texture','worst_perimeter','worst_area','worst_smoothness'
 ↪points','worst_symmetry','worst_fractal dimension']
```

[216]:

```
data.to_csv('labeledData.csv',header=headers,index=False)
```

[217]:

```
data=pd.read_csv('labeledData.csv')
data.head()
```

[217]:

```
         id diagnosis  mean_radius  mean_texture  mean_perimeter  mean_area
0    842302         M        17.99         10.38          122.80     1001.0  \
1    842517         M        20.57         17.77          132.90     1326.0
2  84300903         M        19.69         21.25          130.00     1203.0
3  84348301         M        11.42         20.38           77.58      386.1
4  84358402         M        20.29         14.34          135.10     1297.0

   mean_smoothness  mean_compactness  mean_concavity  mean_concave points
0          0.11840           0.27760          0.3001              0.14710  \
1          0.08474           0.07864          0.0869              0.07017
2          0.10960           0.15990          0.1974              0.12790
3          0.14250           0.28390          0.2414              0.10520
4          0.10030           0.13280          0.1980              0.10430

   …  worst_radius  worst_texture  worst_perimeter  worst_area
0  …         25.38          17.33           184.60      2019.0  \
1  …         24.99          23.41           158.80      1956.0
2  …         23.57          25.53           152.50      1709.0
3  …         14.91          26.50            98.87       567.7
4  …         22.54          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
0          0.4601                  0.11890
1          0.2750                  0.08902
2          0.3613                  0.08758
3          0.6638                  0.17300
```

```
4          0.2364                  0.07678

[5 rows x 32 columns]
```

[218]:

```
def diag(z):
    if z=='M':
        return 1
    else:
        return 0
z=data['diagnosis'].apply(diag)
data.diagnosis=z
```

[221]:

```
df=pd.DataFrame(data=data)
df=df.drop('id',axis=1)
x=df.drop('diagnosis',axis=1)
y=df['diagnosis']
```

[220]:

```
x_scaled=x.apply(zscore)
x_scaled.describe()
```

[220]:

```
        mean_radius  mean_texture  mean_perimeter     mean_area
count  5.690000e+02  5.690000e+02    5.690000e+02  5.690000e+02  \
mean  -1.373633e-16  6.868164e-17   -1.248757e-16 -2.185325e-16
std    1.000880e+00  1.000880e+00    1.000880e+00  1.000880e+00
min   -2.029648e+00 -2.229249e+00   -1.984504e+00 -1.454443e+00
25%   -6.893853e-01 -7.259631e-01   -6.919555e-01 -6.671955e-01
50%   -2.150816e-01 -1.046362e-01   -2.359800e-01 -2.951869e-01
75%    4.693926e-01  5.841756e-01    4.996769e-01  3.635073e-01
max    3.971288e+00  4.651889e+00    3.976130e+00  5.250529e+00

       mean_smoothness  mean_compactness  mean_concavity  mean_concave points
count     5.690000e+02      5.690000e+02    5.690000e+02         5.690000e+02  \
mean     -8.366672e-16      1.873136e-16    4.995028e-17        -4.995028e-17
std       1.000880e+00      1.000880e+00    1.000880e+00         1.000880e+00
min      -3.112085e+00     -1.610136e+00   -1.114873e+00        -1.261820e+00
25%      -7.109628e-01     -7.470860e-01   -7.437479e-01        -7.379438e-01
50%      -3.489108e-02     -2.219405e-01   -3.422399e-01        -3.977212e-01
75%       6.361990e-01      4.938569e-01    5.260619e-01         6.469351e-01
max       4.770911e+00      4.568425e+00    4.243589e+00         3.927930e+00

       mean_symmetry  mean_fractal dimension  …  worst_radius
count   5.690000e+02            5.690000e+02  …  5.690000e+02  \
mean    1.748260e-16            4.745277e-16  … -8.241796e-16
std     1.000880e+00            1.000880e+00  …  1.000880e+00
min    -2.744117e+00           -1.819865e+00  … -1.726901e+00
25%    -7.032397e-01           -7.226392e-01  … -6.749213e-01
50%    -7.162650e-02           -1.782793e-01  … -2.690395e-01
```

```
75%     5.307792e-01            4.709834e-01  …  5.220158e-01
max     4.484751e+00            4.910919e+00  …  4.094189e+00

       worst_texture  worst_perimeter  worst_area  worst_smoothness  \
count   5.690000e+02     5.690000e+02  569.000000      5.690000e+02
mean    1.248757e-17    -3.746271e-16    0.000000     -2.372638e-16
std     1.000880e+00     1.000880e+00    1.000880      1.000880e+00
min    -2.223994e+00    -1.693361e+00   -1.222423     -2.682695e+00
25%    -7.486293e-01    -6.895783e-01   -0.642136     -6.912304e-01
50%    -4.351564e-02    -2.859802e-01   -0.341181     -4.684277e-02
75%     6.583411e-01     5.402790e-01    0.357589      5.975448e-01
max     3.885905e+00     4.287337e+00    5.930172      3.955374e+00

       worst_compactness  worst_concavity  worst_concave points  \
count       5.690000e+02     5.690000e+02          5.690000e+02
mean       -3.371644e-16     7.492542e-17          2.247763e-16
std         1.000880e+00     1.000880e+00          1.000880e+00
min        -1.443878e+00    -1.305831e+00         -1.745063e+00
25%        -6.810833e-01    -7.565142e-01         -7.563999e-01
50%        -2.695009e-01    -2.182321e-01         -2.234689e-01
75%         5.396688e-01     5.311411e-01          7.125100e-01
max         5.112877e+00     4.700669e+00          2.685877e+00

       worst_symmetry  worst_fractal dimension
count    5.690000e+02             5.690000e+02
mean     2.622390e-16            -5.744282e-16
std      1.000880e+00             1.000880e+00
min     -2.160960e+00            -1.601839e+00
25%     -6.418637e-01            -6.919118e-01
50%     -1.274095e-01            -2.164441e-01
75%      4.501382e-01             4.507624e-01
max      6.046041e+00             6.846856e+00

[8 rows x 30 columns]
```

```
[229]: from sklearn.model_selection import train_test_split
       random_state=42
       x_train,x_test,y_train,y_test=train_test_split(x_scaled,y,test_size=0.
        ↪3,random_state=random_state)
```

```
[233]: from sklearn.svm import SVC
       from sklearn.model_selection import KFold, GridSearchCV
       from sklearn.metrics import fbeta_score, make_scorer
       ftwo_scorer = make_scorer(fbeta_score, beta=2)

       c_values = np.arange(0, 1, 0.001)
       kernel_values = ['linear', 'poly', 'rbf', 'sigmoid']
```

```
param_grid = dict(C=c_values, kernel=kernel_values)
model = SVC()
kfold = KFold(n_splits=5)
grid = GridSearchCV(estimator=model, param_grid=param_grid,␣
 ↪scoring=ftwo_scorer, cv=kfold)
grid_result = grid.fit(x_train, y_train)
print("Best: %f using %s" % (grid_result.best_score_, grid_result.best_params_))
```

```
Best: 0.959207 using {'C': 0.08600000000000001, 'kernel': 'linear'}
```

[234]:
```
best_model = grid_result.best_estimator_
best_model.fit(x_train, y_train)
y_pred = best_model.predict(x_test)
```

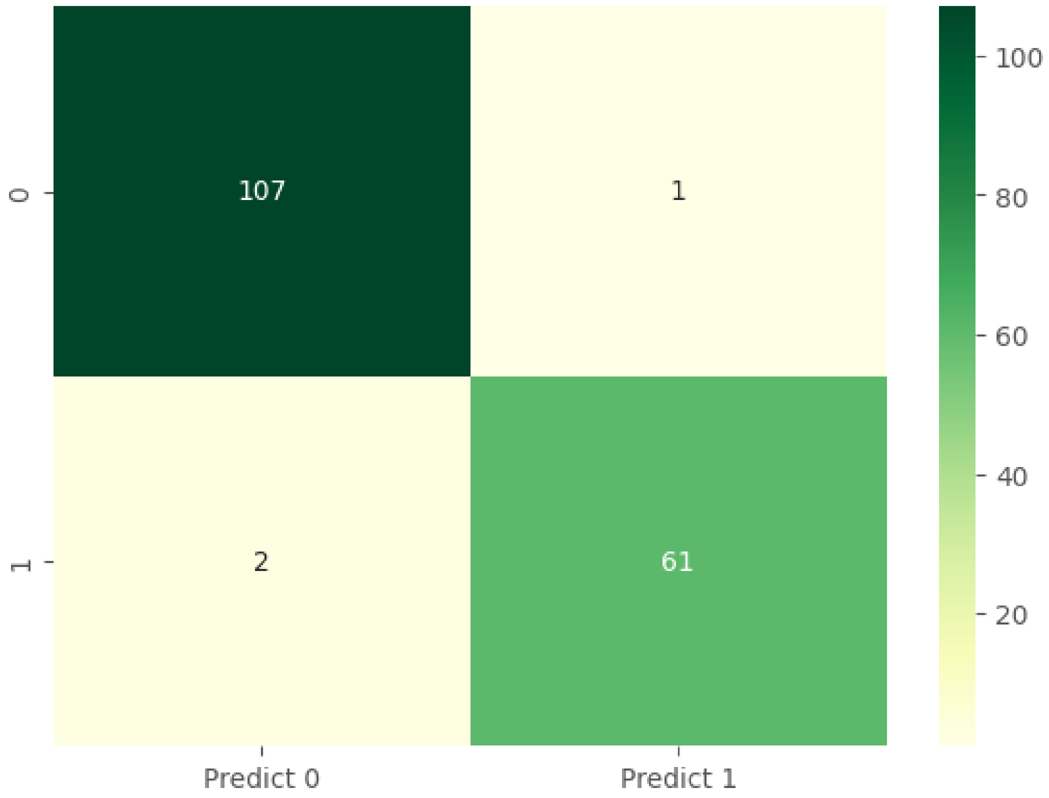
[235]:
```
best_model.score(x_test,y_test)
```

[235]:
```
0.9824561403508771
```

[236]:
```
print('Confusion Matrix')
cm=metrics.confusion_matrix(y_test,y_pred,labels=[0,1])
df_cm=pd.DataFrame(cm,index=[i for i in [0,1]],
                   columns=[i for i in ['Predict 0','Predict 1']])
plt.figure(figsize=(7,5))
sns.heatmap(df_cm,annot=True,fmt='.5g',cmap='YlGn')
```

```
Confusion Matrix
```

[236]:
```
<Axes: >
```



```
[237]: false_negatives=np.logical_and(y_test!=y_pred,y_pred==0)
       x_test[false_negatives]
```

```
[237]:     mean_radius  mean_texture  mean_perimeter  mean_area  mean_smoothness  \
       73     -0.092956     -0.814392       -0.063393  -0.201331         0.308838
       255    -0.047513     -0.521181       -0.022203  -0.149284         0.942210

            mean_compactness  mean_concavity  mean_concave points  mean_symmetry  \
       73           0.448373       -0.136966             0.045677      -0.546249
       255          0.446478        0.114133             0.091333       0.351883

            mean_fractal dimension  …  worst_radius  worst_texture  \
       73                 0.405774  …      0.062293      -0.784455
       255               -0.212302  …      0.025018      -0.587414

            worst_perimeter  worst_area  worst_smoothness  worst_compactness  \
       73          0.090513   -0.119860          0.382749           0.635726
       255         0.024984   -0.095952          0.825491           0.457607

            worst_concavity  worst_concave points  worst_symmetry  \
       73          0.027401              0.360776       -0.504352
```

```
255         0.233695              0.347072        0.270565

     worst_fractal dimension
73                  1.055903
255                -0.242489

[2 rows x 30 columns]
```