SVM

October 25, 2022

[39]:
```
import numpy as np
import pandas as pd
```

[40]:
```
data= pd.read_csv("voice-classification.csv")
```

[41]:
```
data.head()
```

[41]:

	meanfreq	sd	median	Q25	Q75	IQR	skew	\
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462	
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285	
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155	
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174	

	kurt	sp.ent	sfm	…	centroid	meanfun	minfun	\
0	274.402906	0.893369	0.491918	…	0.059781	0.084279	0.015702	
1	634.613855	0.892193	0.513724	…	0.066009	0.107937	0.015826	
2	1024.927705	0.846389	0.478905	…	0.077316	0.098706	0.015656	
3	4.177296	0.963322	0.727232	…	0.151228	0.088965	0.017798	
4	4.333713	0.971955	0.783568	…	0.135120	0.106398	0.016931	

	maxfun	meandom	mindom	maxdom	dfrange	modindx	label
0	0.275862	0.007812	0.007812	0.007812	0.000000	0.000000	male
1	0.250000	0.009014	0.007812	0.054688	0.046875	0.052632	male
2	0.271186	0.007990	0.007812	0.015625	0.007812	0.046512	male
3	0.250000	0.201497	0.007812	0.562500	0.554688	0.247119	male
4	0.266667	0.712812	0.007812	5.484375	5.476562	0.208274	male

[5 rows x 21 columns]

[42]:
```
data.isna().sum()
```

[42]:
```
meanfreq    0
sd          0
median      0
Q25         0
Q75         0
```

```
IQR         0
skew        0
kurt        0
sp.ent      0
sfm         0
mode        0
centroid    0
meanfun     0
minfun      0
maxfun      0
meandom     0
mindom      0
maxdom      0
dfrange     0
modindx     0
label       0
dtype: int64
```

[43]:
```
data.shape
```

[43]:
```
(3168, 21)
```

[44]:
```
data.size
```

[44]:
```
66528
```

[45]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3168 entries, 0 to 3167
Data columns (total 21 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   meanfreq  3168 non-null   float64
 1   sd        3168 non-null   float64
 2   median    3168 non-null   float64
 3   Q25       3168 non-null   float64
 4   Q75       3168 non-null   float64
 5   IQR       3168 non-null   float64
 6   skew      3168 non-null   float64
 7   kurt      3168 non-null   float64
 8   sp.ent    3168 non-null   float64
 9   sfm       3168 non-null   float64
 10  mode      3168 non-null   float64
 11  centroid  3168 non-null   float64
 12  meanfun   3168 non-null   float64
 13  minfun    3168 non-null   float64
```

```
 14  maxfun    3168 non-null   float64
 15  meandom   3168 non-null   float64
 16  mindom    3168 non-null   float64
 17  maxdom    3168 non-null   float64
 18  dfrange   3168 non-null   float64
 19  modindx   3168 non-null   float64
 20  label     3168 non-null   object
dtypes: float64(20), object(1)
memory usage: 519.9+ KB
```

[46]:
```
x=data.drop("label", axis=1)
```

[47]:
```
x.head()
```

[47]:

	meanfreq	sd	median	Q25	Q75	IQR	skew	\
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462	
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285	
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155	
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174	

	kurt	sp.ent	sfm	mode	centroid	meanfun	minfun	\
0	274.402906	0.893369	0.491918	0.000000	0.059781	0.084279	0.015702	
1	634.613855	0.892193	0.513724	0.000000	0.066009	0.107937	0.015826	
2	1024.927705	0.846389	0.478905	0.000000	0.077316	0.098706	0.015656	
3	4.177296	0.963322	0.727232	0.083878	0.151228	0.088965	0.017798	
4	4.333713	0.971955	0.783568	0.104261	0.135120	0.106398	0.016931	

	maxfun	meandom	mindom	maxdom	dfrange	modindx
0	0.275862	0.007812	0.007812	0.007812	0.000000	0.000000
1	0.250000	0.009014	0.007812	0.054688	0.046875	0.052632
2	0.271186	0.007990	0.007812	0.015625	0.007812	0.046512
3	0.250000	0.201497	0.007812	0.562500	0.554688	0.247119
4	0.266667	0.712812	0.007812	5.484375	5.476562	0.208274

[48]:
```
y=data["label"]
```

[49]:
```
y
```

[49]:
```
0         male
1         male
2         male
3         male
4         male
         ...
3163    female
3164    female
```

```
3165    female
3166    female
3167    female
Name: label, Length: 3168, dtype: object
```

[50]:
```
from sklearn.preprocessing import StandardScaler
```

[51]:
```
scaler= StandardScaler()
```

[52]:
```
scaler.fit(x)
x = scaler.transform(x)
```

[55]:
```
from sklearn.preprocessing import LabelEncoder
```

[56]:
```
label= LabelEncoder()
```

[57]:
```
y= label.fit_transform(y)
```

[58]:
```
y
```

[58]:
```
array([1, 1, 1, …, 0, 0, 0])
```

[59]:
```
from sklearn.model_selection import train_test_split
```

[60]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=2)
```

[63]:
```
print(x.shape, x_train.shape, x_test.shape)
```

```
(3168, 20) (2376, 20) (792, 20)
```

[66]:
```
from sklearn.svm import SVC
```

[68]:
```
svc= SVC()
```

[69]:
```
svc.fit(x_train,y_train)
```

[69]:
```
SVC()
```

[70]:
```
from sklearn.metrics import accuracy_score, confusion_matrix,␣
↪classification_report
```

[72]:
```
ypred=svc.predict(x_test)
```

[73]:
```
accuracy_score(y_test,ypred)
```

[73]:
```
0.976010101010101
```

[74]:
```
print(classification_report(y_test,ypred))
```

```
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       430
           1       0.97      0.98      0.97       362

    accuracy                           0.98       792
   macro avg       0.98      0.98      0.98       792
weighted avg       0.98      0.98      0.98       792
```

[75]:
```
confusion_matrix(y_test,ypred)
```

[75]:
```
array([[420,  10],
       [  9, 353]])
```

[ ]:
```

```