ML - Lab Assignment 5
SVM

[85]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.svm import SVC
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, classification_report
```

[86]:
```
df= pd.read_csv("thyroidDF.csv")
```

[87]:
```
df.head()
```

[87]:
```
   age sex on_thyroxine query_on_thyroxine on_antithyroid_meds sick pregnant  \
0   29   F            f                  f                   f    f        f
1   29   F            f                  f                   f    f        f
2   41   F            f                  f                   f    f        f
3   36   F            f                  f                   f    f        f
4   32   F            f                  f                   f    f        f

  thyroid_surgery I131_treatment query_hypothyroid  …    TT4 T4U_measured  \
0               f              f                 t  …    NaN            f
1               f              f                 f  …  128.0            f
2               f              f                 f  …    NaN            f
3               f              f                 f  …    NaN            f
4               f              f                 f  …    NaN            f

  T4U FTI_measured FTI TBG_measured   TBG  referral_source target  patient_id
0 NaN            f NaN            f   NaN            other      -   840801013
1 NaN            f NaN            f   NaN            other      -   840801014
2 NaN            f NaN            t  11.0            other      -   840801042
3 NaN            f NaN            t  26.0            other      -   840803046
4 NaN            f NaN            t  36.0            other      S   840803047

[5 rows x 31 columns]
```

[88]:
```
df['target'].sort_values().unique()
```

```
[88]: array(['-', 'A', 'AK', 'B', 'C', 'C|I', 'D', 'D|R', 'E', 'F', 'FK', 'G',
             'GI', 'GK', 'GKJ', 'H|K', 'I', 'J', 'K', 'KJ', 'L', 'LJ', 'M',
             'MI', 'MK', 'N', 'O', 'OI', 'P', 'Q', 'R', 'S'], dtype=object)
```

```
[89]: mapping = {'-':'Negative',
                 'A':'Hyperthyroid','AK':'Hyperthyroid','B':'Hyperthyroid','C':
       ↪'Hyperthyroid','C|I':'Hyperthyroid','D':'Hyperthyroid','D|R':'Hyperthyroid',
                 'E':'Hypothyroid','F':'Hypothyroid','FK':'Hypothyroid','G':
       ↪'Hypothyroid','GK':'Hypothyroid','GI':'Hypothyroid','GKJ':
       ↪'Hypothyroid','H|K':'Hypothyroid',
                 'I':'Hypothyroid','J':'Hypothyroid','K':'Hypothyroid','KJ':
       ↪'Hypothyroid','L':'Hypothyroid','LJ':'Hypothyroid','M':'Hypothyroid','MI':
       ↪'Hypothyroid',
                 'MK':'Hypothyroid','N':'Hypothyroid','O':'Hypothyroid','OI':
       ↪'Hypothyroid','P':'Hypothyroid','Q':'Hypothyroid','R':'Hypothyroid','S':
       ↪'Hypothyroid',
      }
```

```
[90]: df['target'] = df['target'].map(mapping)
```

```
[91]: df['target'].unique()
```

```
[91]: array(['Negative', 'Hypothyroid', 'Hyperthyroid'], dtype=object)
```

```
[92]: df.head()
```

```
[92]:    age sex on_thyroxine query_on_thyroxine on_antithyroid_meds sick pregnant  \
      0   29   F            f                  f                   f    f        f
      1   29   F            f                  f                   f    f        f
      2   41   F            f                  f                   f    f        f
      3   36   F            f                  f                   f    f        f
      4   32   F            f                  f                   f    f        f

        thyroid_surgery I131_treatment query_hypothyroid  …    TT4 T4U_measured  \
      0               f              f                 t  …    NaN            f
      1               f              f                 f  …  128.0            f
      2               f              f                 f  …    NaN            f
      3               f              f                 f  …    NaN            f
      4               f              f                 f  …    NaN            f

        T4U FTI_measured FTI TBG_measured   TBG referral_source       target  \
      0 NaN            f NaN            f   NaN           other     Negative
      1 NaN            f NaN            f   NaN           other     Negative
      2 NaN            f NaN            t  11.0           other     Negative
      3 NaN            f NaN            t  26.0           other     Negative
      4 NaN            f NaN            t  36.0           other  Hypothyroid
```

```
   patient_id
0   840801013
1   840801014
2   840801042
3   840803046
4   840803047

[5 rows x 31 columns]
```

[93]:

```
df.isna().sum()
```

[93]:

```
age                       0
sex                     307
on_thyroxine              0
query_on_thyroxine        0
on_antithyroid_meds       0
sick                      0
pregnant                  0
thyroid_surgery           0
I131_treatment            0
query_hypothyroid         0
query_hyperthyroid        0
lithium                   0
goitre                    0
tumor                     0
hypopituitary             0
psych                     0
TSH_measured              0
TSH                     842
T3_measured               0
T3                     2604
TT4_measured              0
TT4                     442
T4U_measured              0
T4U                     809
FTI_measured              0
FTI                     802
TBG_measured              0
TBG                    8823
referral_source           0
target                    0
patient_id                0
dtype: int64
```

[94]:

```
df = df.drop(['sex','referral_source','patient_id'],axis=1)
```

[95]:

```
df.dropna(inplace=True)
```

```
[96]: df.head()
```

```
[96]:       age on_thyroxine query_on_thyroxine on_antithyroid_meds sick pregnant  \
      167    40            f                  f                   f    f        f
      5256   35            f                  f                   f    f        f
      6044   77            f                  f                   f    f        f
      6045   73            f                  f                   f    f        f
      6681   78            f                  f                   f    f        f

           thyroid_surgery I131_treatment query_hypothyroid query_hyperthyroid  ...  \
      167                f              f                 f                  f  ...
      5256               t              f                 f                  f  ...
      6044               f              f                 f                  f  ...
      6045               f              f                 f                  f  ...
      6681               f              f                 f                  f  ...

             T3 TT4_measured    TT4 T4U_measured   T4U FTI_measured    FTI  \
      167   0.4            t    3.9            t  0.83            t    5.0
      5256  1.9            t   73.0            t  1.16            t   63.0
      6044  1.8            t  120.0            t  0.96            t  124.0
      6045  1.6            t   89.0            t  0.74            t  119.0
      6681  2.2            t  103.0            t  0.90            t  114.0

           TBG_measured   TBG       target
      167             t  28.0  Hypothyroid
      5256            t  37.0     Negative
      6044            t  45.0     Negative
      6045            t  24.0     Negative
      6681            t  30.0  Hypothyroid

      [5 rows x 28 columns]
```

```
[97]: lbl_encoder = LabelEncoder()
      df['target']=lbl_encoder.fit_transform(df['target'])
```

```
[98]: df=df.replace('f',0)
      df=df.replace('t',1)
```

```
[99]: df.head()
```

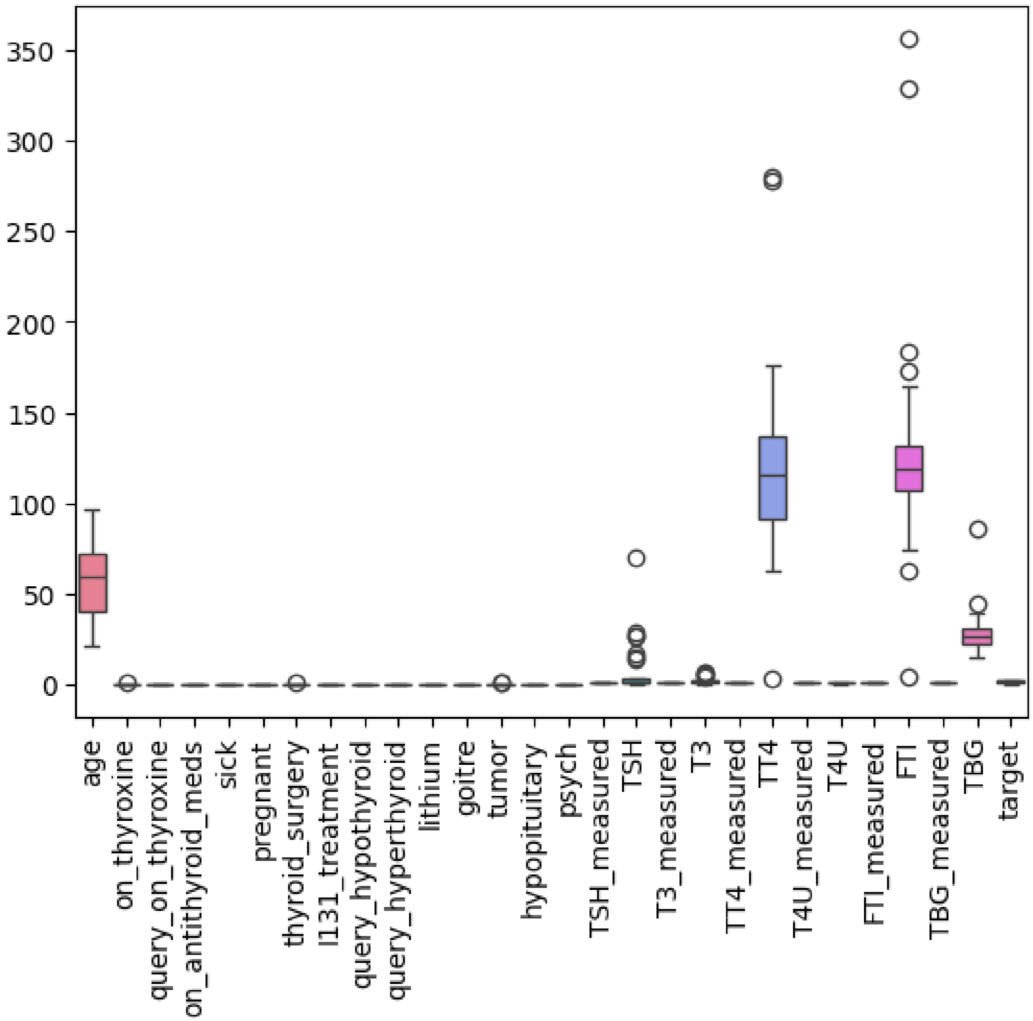
```
[99]:       age  on_thyroxine  query_on_thyroxine  on_antithyroid_meds  sick  \
      167    40             0                   0                    0     0
      5256   35             0                   0                    0     0
      6044   77             0                   0                    0     0
      6045   73             0                   0                    0     0
      6681   78             0                   0                    0     0
      pregnant  thyroid_surgery  I131_treatment  query_hypothyroid  \
167          0                0               0                  0
5256         0                1               0                  0
6044         0                0               0                  0
6045         0                0               0                  0
6681         0                0               0                  0

      query_hyperthyroid  …   T3  TT4_measured    TT4  T4U_measured   T4U  \
167                    0  …  0.4             1    3.9             1  0.83
5256                   0  …  1.9             1   73.0             1  1.16
6044                   0  …  1.8             1  120.0             1  0.96
6045                   0  …  1.6             1   89.0             1  0.74
6681                   0  …  2.2             1  103.0             1  0.90

      FTI_measured    FTI  TBG_measured   TBG  target
167              1    5.0             1  28.0       1
5256             1   63.0             1  37.0       2
6044             1  124.0             1  45.0       2
6045             1  119.0             1  24.0       2
6681             1  114.0             1  30.0       1

[5 rows x 28 columns]
```

[100]:

```python
sns.boxplot(df)
plt.xticks(rotation=90)
```

[100]:

```
([0,
  1,
  2,
  3,
  4,
  5,
  6,
  7,
  8,
  9,
  10,
  11,
  12,
  13,
  14,
  15,
  16,
  17,
  18,
  19,
  20,
 21,
 22,
 23,
 24,
 25,
 26,
 27],
[Text(0, 0, 'age'),
 Text(1, 0, 'on_thyroxine'),
 Text(2, 0, 'query_on_thyroxine'),
 Text(3, 0, 'on_antithyroid_meds'),
 Text(4, 0, 'sick'),
 Text(5, 0, 'pregnant'),
 Text(6, 0, 'thyroid_surgery'),
 Text(7, 0, 'I131_treatment'),
 Text(8, 0, 'query_hypothyroid'),
 Text(9, 0, 'query_hyperthyroid'),
 Text(10, 0, 'lithium'),
 Text(11, 0, 'goitre'),
 Text(12, 0, 'tumor'),
 Text(13, 0, 'hypopituitary'),
 Text(14, 0, 'psych'),
 Text(15, 0, 'TSH_measured'),
 Text(16, 0, 'TSH'),
 Text(17, 0, 'T3_measured'),
 Text(18, 0, 'T3'),
 Text(19, 0, 'TT4_measured'),
 Text(20, 0, 'TT4'),
 Text(21, 0, 'T4U_measured'),
 Text(22, 0, 'T4U'),
 Text(23, 0, 'FTI_measured'),
 Text(24, 0, 'FTI'),
 Text(25, 0, 'TBG_measured'),
 Text(26, 0, 'TBG'),
 Text(27, 0, 'target')])
```



[101]:

```python
def outliers (df, ft):

    Q1=df[ft].quantile (0.25)
    Q3=df[ft].quantile (0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 -  1.5*IQR
    upper_bound = Q3 + 1.5* IQR
    ls = df.index[ (df[ft] < lower_bound) | (df[ft] > upper_bound) ]
    return ls
```

[102]:

```python
index_list = []
for feature in ["TSH", "T3", "TT4", "FTI", "TBG"]:
    index_list.extend(outliers(df, feature))
```

[103]:

```
def remove(df, ls):
    ls = sorted(set(ls))
    df = df.drop(ls)
    return df
```

[104]:

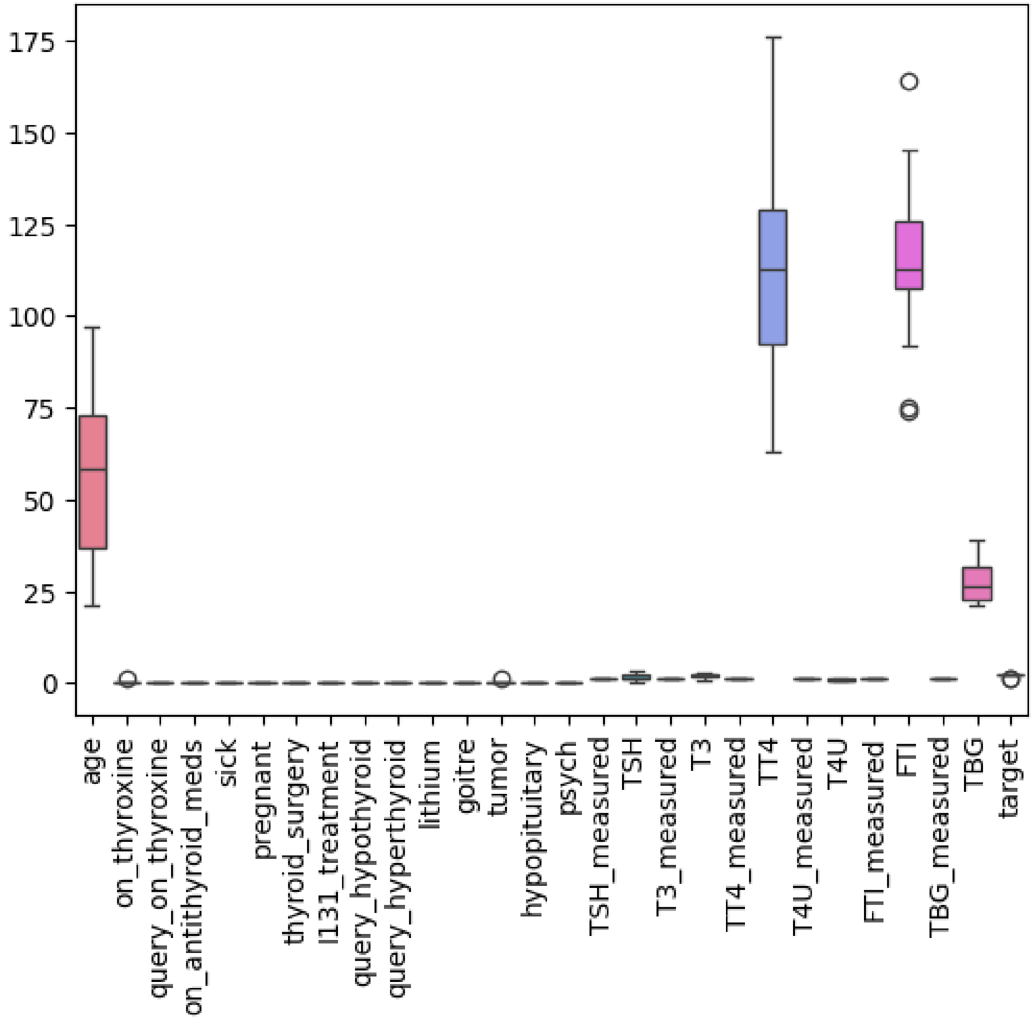
```
df_cleaned = remove(df, index_list)
```

[105]:

```
sns.boxplot(df_cleaned)
plt.xticks(rotation=90)
```

[105]:

```
([0,
  1,
  2,
  3,
  4,
  5,
  6,
  7,
  8,
  9,
  10,
  11,
  12,
  13,
  14,
  15,
  16,
  17,
  18,
  19,
  20,
  21,
  22,
  23,
  24,
  25,
  26,
  27],
 [Text(0, 0, 'age'),
  Text(1, 0, 'on_thyroxine'),
  Text(2, 0, 'query_on_thyroxine'),
  Text(3, 0, 'on_antithyroid_meds'),
  Text(4, 0, 'sick'),
  Text(5, 0, 'pregnant'),
  Text(6, 0, 'thyroid_surgery'),
  Text(7, 0, 'I131_treatment'),
  Text(8, 0, 'query_hypothyroid'),
  Text(9, 0, 'query_hyperthyroid'),
  Text(10, 0, 'lithium'),
  Text(11, 0, 'goitre'),
  Text(12, 0, 'tumor'),
  Text(13, 0, 'hypopituitary'),
  Text(14, 0, 'psych'),
  Text(15, 0, 'TSH_measured'),
  Text(16, 0, 'TSH'),
  Text(17, 0, 'T3_measured'),
  Text(18, 0, 'T3'),
  Text(19, 0, 'TT4_measured'),
  Text(20, 0, 'TT4'),
  Text(21, 0, 'T4U_measured'),
  Text(22, 0, 'T4U'),
  Text(23, 0, 'FTI_measured'),
  Text(24, 0, 'FTI'),
  Text(25, 0, 'TBG_measured'),
  Text(26, 0, 'TBG'),
  Text(27, 0, 'target')])
```



[106]:
```
X = df_cleaned.drop("target", axis = 1)
y = df_cleaned["target"]
```

[107]:
```
Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=0)
```

[108]:
```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

X_train_std = scaler.fit_transform(Xtrain)# scaling of mean of the 80% training␣
 ↪set
X_test_std = scaler.transform(Xtest)
```

[109]:
```
svm=SVC(kernel='linear',decision_function_shape='ovo', verbose=True)
```

```
[110]: svm.fit(X_train_std, ytrain)
```

```
[LibSVM]
```

```
[110]: SVC(decision_function_shape='ovo', kernel='linear', verbose=True)
```

```
[111]: svm.score(X_train_std, ytrain)
```

```
[111]: 1.0
```

```
[112]: svm.score(X_test_std, ytest)
```

```
[112]: 0.75
```