

ECT_HW5

2019

第一大題

請用python依照步驟對voice.csv 進行SVM分析，過程中對所有重要程式步驟進行截圖並加以說明，越詳盡越好。

第一大題-題目

使用 Python:(60%)↵

- (a) 請檢查資料集是否有空值, 如有空值即去掉該筆資料↵
- (b) 將最後一個屬性值"label"切分為 Target, 其餘屬性切分為 Feature↵
- (c) 將 Target 進行 encoded, 用 LabelEncoder 將 male 轉為 1, female 轉為 0↵
- (d) 將 Feature 用 sklearn.preprocessing 的 StandardScaler 進行標準化↵
- (e) 切分資料集與測試集, 設 test size=0.33, random state=1↵
- (f) 最後, 使用 sklearn.svm 裡的 SVC 進行分析, kernel 設為 'linear', 並印出模型最終的準確度↵

第一大題-解答

```
In [25]: df.isnull().sum()
```

經檢查，本資料集無空值。

```
Out[25]: 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
```

第一大題-解答

將資料集切分為Feature和Target，並將Target encoder

```
In [28]: X=df.iloc[:, :-1]  
X.head()
```

Out[28]:

	meanfreq	sd	median	Q25	Q75	IQR	skew	kurt	sp.ent	sfm	mode	centroid	meanfun	minfun	maxfun
0	0.059781	0.064241	0.032027	0.015071	0.090193	0.075122	12.863462	274.402906	0.893369	0.491918	0.000000	0.059781	0.084279	0.015702	0.275862
1	0.066009	0.067310	0.040229	0.019414	0.092666	0.073252	22.423285	634.613855	0.892193	0.513724	0.000000	0.066009	0.107937	0.015826	0.250000
2	0.077316	0.083829	0.036718	0.008701	0.131908	0.123207	30.757155	1024.927705	0.846389	0.478905	0.000000	0.077316	0.098706	0.015656	0.271186
3	0.151228	0.072111	0.158011	0.096582	0.207955	0.111374	1.232831	4.177296	0.963322	0.727232	0.083878	0.151228	0.088965	0.017798	0.250000
4	0.135120	0.079146	0.124656	0.078720	0.206045	0.127325	1.101174	4.333713	0.971955	0.783568	0.104261	0.135120	0.106398	0.016931	0.266667

```
In [29]: from sklearn.preprocessing import LabelEncoder  
y=df.iloc[:, -1]  
  
# male -> 1  
# female -> 0  
  
gender_encoder = LabelEncoder()  
y = gender_encoder.fit_transform(y)  
y
```

Out[29]: array([1, 1, 1, ..., 0, 0, 0])

第一大題-解答

將Feature標準化

```
In [37]: from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
scaler.fit(X)  
X = scaler.transform(X)
```

切分訓練集與測試集

```
In [31]: from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
```

第一大題-解答

SVC分析

```
In [38]: from sklearn.svm import SVC
          from sklearn import metrics
          svc=SVC(kernel='linear')
          svc.fit(X_train,y_train)
          y_pred=svc.predict(X_test)
          print('Accuracy Score:')
          print(metrics.accuracy_score(y_test,y_pred))
```

Accuracy Score:
0.9789674952198852

第一大題-題目

使用 Weka 軟體:(40%)↵

- (a) 使用 weka 裡的 SMO function 對 voice.csv 進行 SVM 分析, kernel 設為 'linear', Percentage spilt 設為 66%, 截圖並附上過程及準確率↵

第一大題-解答

Kernel used:

Linear Kernel: $K(x,y) = \langle x,y \rangle$

=== Summary ===

Correctly Classified Instances	1054	97.8644 %
Incorrectly Classified Instances	23	2.1356 %
Kappa statistic	0.9573	
Mean absolute error	0.0214	
Root mean squared error	0.1461	
Relative absolute error	4.2706 %	
Root relative squared error	29.2224 %	
Total Number of Instances	1077	