HW3: SVM

Prof. Chia-Yu Lin

Yuan Ze University

2022 Spring

把訓練好的SVM模型用來預測

- 課堂中已經完成SVM模型
- 可利用test.csv取得測試資料的欄位

```
1  #取得測試資料的欄位
2  testset = testset[testset.columns[:449]]
```

SVM預測

```
1 svm.predict(testset)
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4,
       4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
       4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
       5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
       5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
       5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
       6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
       6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,
       7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7,
       7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7])
```

- 但是老師的test.csv沒有正確答案，無法對答案……
- 所以……..
- 同學可以從訓練資料隨機抓幾筆出來測試預測準確度

HW3-1:
從訓練資料隨機抓幾筆做測試 (1/2)

- 讀取訓練資料
- 確認訓練資料的列數
- 設定random範圍，random 10個整數

```
資料的形式:(1741, 450)
[1062  236 1097  906  716 1670  848  961  145  130]
```

HW3-1:
從訓練資料隨機抓幾筆做測試 (2/2)

- 將random的結果當成index，取出列

	col1	col2	col3	col4	col5	col6	col7	col8	col9	col10	...	col440	col441	col442	col443	col444	col445	col446	col447	col448	col449
1062	70.1	70.1	70.1	70.0	70.2	70.2	70.4	70.5	70.7	70.8	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
236	78.6	78.6	78.6	78.6	79.2	79.9	80.5	81.3	82.2	83.6	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1097	81.4	81.4	81.4	81.5	81.8	82.0	82.3	82.6	82.9	83.1	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
906	76.2	76.2	76.2	76.2	76.4	76.7	77.0	77.5	78.2	79.2	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
716	69.7	69.7	69.8	69.8	72.0	71.3	72.4	75.7	75.4	77.2	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1670	70.0	70.0	70.0	70.1	70.4	70.7	70.8	71.0	71.2	71.5	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
848	84.9	84.9	84.9	84.9	84.9	85.1	85.4	86.2	86.6	87.5	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
961	72.3	72.3	72.3	72.3	72.8	73.3	74.2	75.2	76.5	78.2	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
145	81.2	81.2	81.2	81.2	81.4	82.0	83.0	84.1	85.2	86.7	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
130	92.0	92.0	92.0	92.0	93.4	94.2	96.1	98.3	101.0	103.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

10 rows × 449 columns

SVM預測

- 預測

```
1 predicted=svm.predict(dataset_fortest)
2 predicted
```

```
array([5, 1, 5, 4, 4, 7, 4, 4, 3, 0])
```

- 實際答案

```
1 expected=label.values
2 expected
```

```
array([5, 1, 5, 4, 4, 7, 4, 4, 1, 0])
```

預測的準確度

```
print(metrics.classification_report(expected,predicted))
print(metrics.confusion_matrix(expected, predicted))
accuracy = accuracy_score(expected, predicted)
print("Accuracy: %.2f%%" % (accuracy * 100.0))
print('precision:',metrics.precision_score(expected, predicted,average='macro'))
print('recall:',metrics.recall_score(expected, predicted,average='macro'))
print('F1-score:',metrics.f1_score(label,predicted,average='macro'))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       1.00      0.50      0.67         2
           3       0.00      0.00      0.00         0
           4       1.00      1.00      1.00         4
           5       1.00      1.00      1.00         2
           7       1.00      1.00      1.00         1

    accuracy                           0.90        10
   macro avg       0.83      0.75      0.78        10
weighted avg       1.00      0.90      0.93        10

[[1 0 0 0 0 0]
 [0 1 1 0 0 0]
 [0 0 0 0 0 0]
 [0 0 0 4 0 0]
 [0 0 0 0 2 0]
 [0 0 0 0 0 1]]
Accuracy: 90.00%
precision: 0.8333333333333334
recall: 0.75
F1-score: 0.7777777777777777
```

HW3-2

- 利用隨機森林再重新操作上述的步驟

作業繳交方式

- 請按照以下的說明完成” HW3-SVM.ipynb”
- 將SVM模型的預測答案和實際答案截圖，預測準確度截圖

```
predicted=svm.predict(dataset_fortest)
predicted
```

```
array([5, 1, 5, 4, 4, 7, 4, 4, 3, 0])
```

```
expected=label.values
expected
```

```
array([5, 1, 5, 4, 4, 7, 4, 4, 1, 0])
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1
1	1.00	0.50	0.67	2
3	0.00	0.00	0.00	0
4	1.00	1.00	1.00	4
5	1.00	1.00	1.00	2
7	1.00	1.00	1.00	1
accuracy			0.90	10
macro avg	0.83	0.75	0.78	10
weighted avg	1.00	0.90	0.93	10

```
[[1 0 0 0 0 0]
 [0 1 1 0 0 0]
 [0 0 0 0 0 0]
 [0 0 0 4 0 0]
 [0 0 0 0 2 0]
 [0 0 0 0 0 1]]
Accuracy: 90.00%
precision: 0.8333333333333334
recall: 0.75
F1-score: 0.7777777777777777
```

- 將隨機森林模型的預測答案和實際答案截圖，預測準確度截圖
- 將上述的截圖放在word檔，連同HW3-SVM.ipynb一起上傳
- Deadline: 3/22 23:59