人工智慧實務期末報告

鳶尾花分類

學號:408630977

系級:資管四A

姓名:蔡品洋

座號:043

動機

PictureThis（形色）
口袋裡的植物專家

線上植物百科全書和植物識別專家

用手機試用PictureThis App，輕鬆認識上萬種花草樹木！

PictureThis

https://www.picturethisai.com/zh-tw/app

變色鳶尾

Iris versicolor

白色鳶尾花

Iris albicans

鳶尾花

Iris fulva

波斯尼亞鳶尾

Iris reichenbachii

無葉鳶尾

Iris aphylla

The iris dataset is a classic and very easy multi-class classification dataset.

Classes	3
Samples per class	50
Samples total	150
Dimensionality	4
Features	real, positive

Target

```
'target': array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 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, 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, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2]),
'frame': None,
'target_names': array(['setosa', 'versicolor', 'virginica'], dtype='<U10'),
```

山鳶尾　，變色鳶尾　　，維吉尼亞鳶尾

Data

```
from sklearn.datasets import load_iris
data = load_iris()

data
```

```
{'data': array([[5.1, 3.5, 1.4, 0.2],
        [4.9, 3. , 1.4, 0.2],
        [4.7, 3.2, 1.3, 0.2],
        [4.6, 3.1, 1.5, 0.2],
        [5. , 3.6, 1.4, 0.2],
        [5.4, 3.9, 1.7, 0.4],
        [4.6, 3.4, 1.4, 0.3],
        [5. , 3.4, 1.5, 0.2],
```

萼片長度/寬度；花瓣長度/寬度

```
'feature_names': ['sepal length (cm)',
 'sepal width (cm)',
 'petal length (cm)',
 'petal width (cm)'],
```

Data

	sepal le萼片長度	sepal w萼片寬度	petal le花瓣長度	petal w花瓣寬度	t種類
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

方法

- **資料集**
  :https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_iris.html#sklearn.datasets.load_iris
- **目標:**透過四個特徵之間的關聯，找出data跟target的相關性，並分類
- **學習方式:**監督式學習 Supervised learning
- **任務:**分類 Classification
- **資料預處理:**不需要

```
[ ]  df.corr()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
sepal length (cm)	1.000000	-0.117570	0.871754	0.817941	0.782561
sepal width (cm)	-0.117570	1.000000	-0.428440	-0.366126	-0.426658
petal length (cm)	0.871754	-0.428440	1.000000	0.962865	0.949035
petal width (cm)	0.817941	-0.366126	0.962865	1.000000	0.956547
target	0.782561	-0.426658	0.949035	0.956547	1.000000

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12,8))
sns.heatmap(df.astype("float").corr(), cmap="summer", annot=True)
```

鄂片長度　鄂片寬度　花瓣長度　花瓣寬度　target

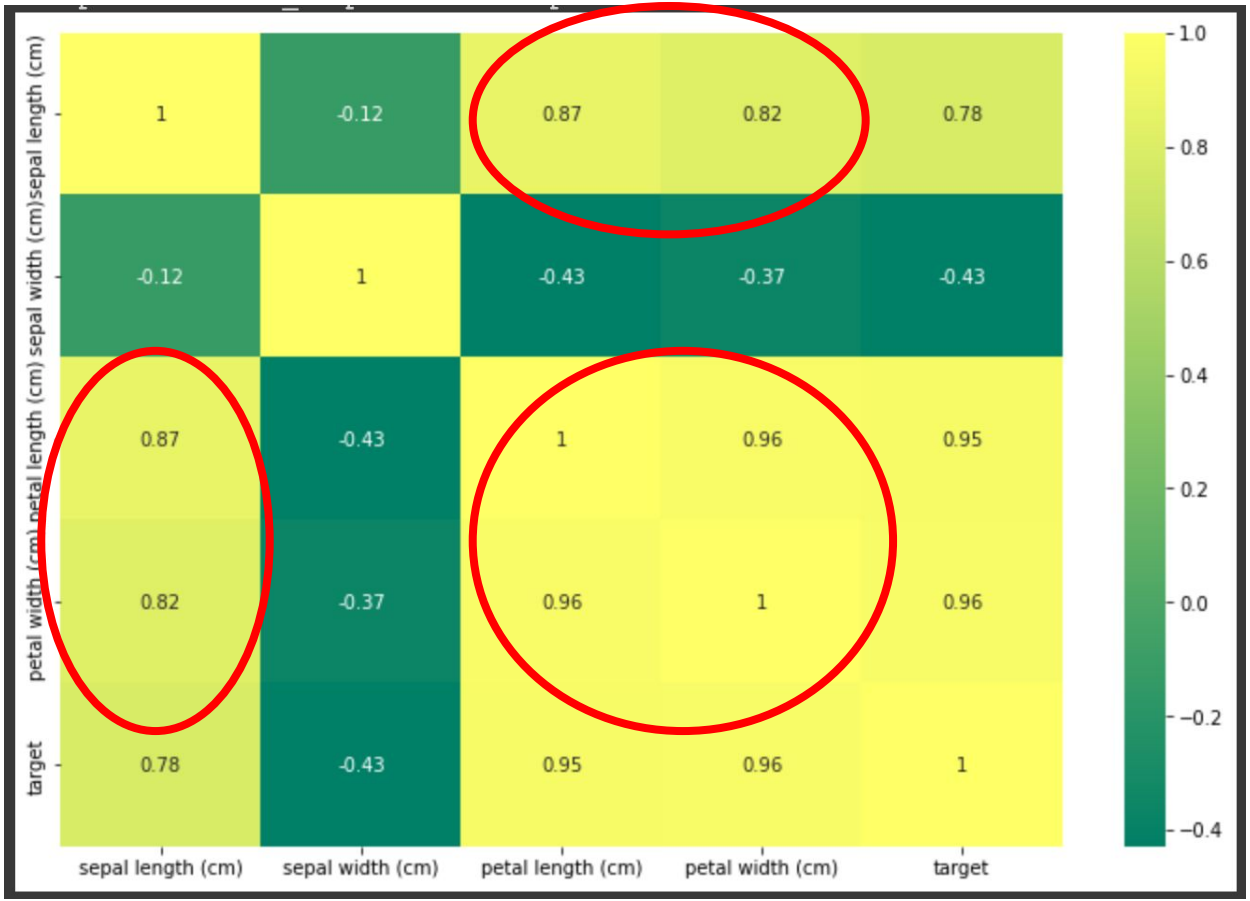
鄂片長度

鄂片寬度

花瓣長度

花瓣寬度

target



```
[127] from sklearn.model_selection import train_test_split

      x_train, x_test, y_train, y_test = train_test_split(df.drop(["target"], axis=1), df["target"], test_size=0.1)
      # print(x_train, x_test, y_train, y_test)
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

90%

訓練資料

10%

測試資料

```
[122] from sklearn.tree import DecisionTreeClassifier

      clf = DecisionTreeClassifier()
      clf.fit(x_train, y_train)
```

```
DecisionTreeClassifier()
```

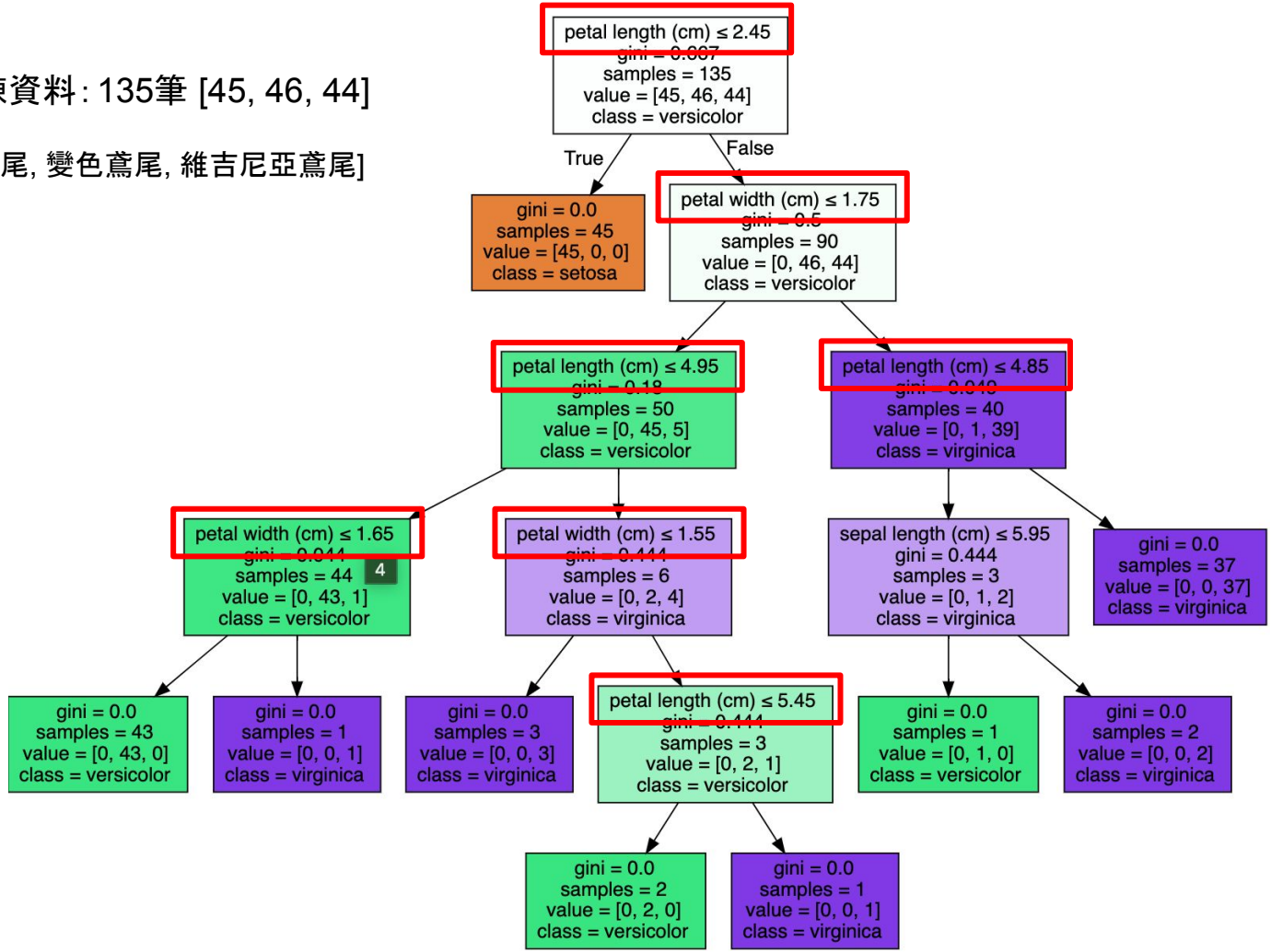
```
[123] from sklearn.tree import export_graphviz
      import graphviz

      g = export_graphviz(clf, out_file=None, feature_names=iris["feature_names"], class_names=iris["target_names"],
                          filled=True, special_characters=True)

      graph = graphviz.Source(g)
      graph
```

- 訓練資料: 135筆 [45, 46, 44]

  [山鳶尾, 變色鳶尾, 維吉尼亞鳶尾]



```
[145] pre = clf.predict(x_test)
      print("target:", list(y_test))
      print("預測結果:", list(pre))
```

```
target: [1, 2, 0, 1, 2, 0, 2, 1, 2, 0, 1, 2, 2, 1, 1]
預測結果: [1, 2, 0, 1, 2, 0, 2, 1, 2, 0, 1, 2, 1, 2, 1]
```

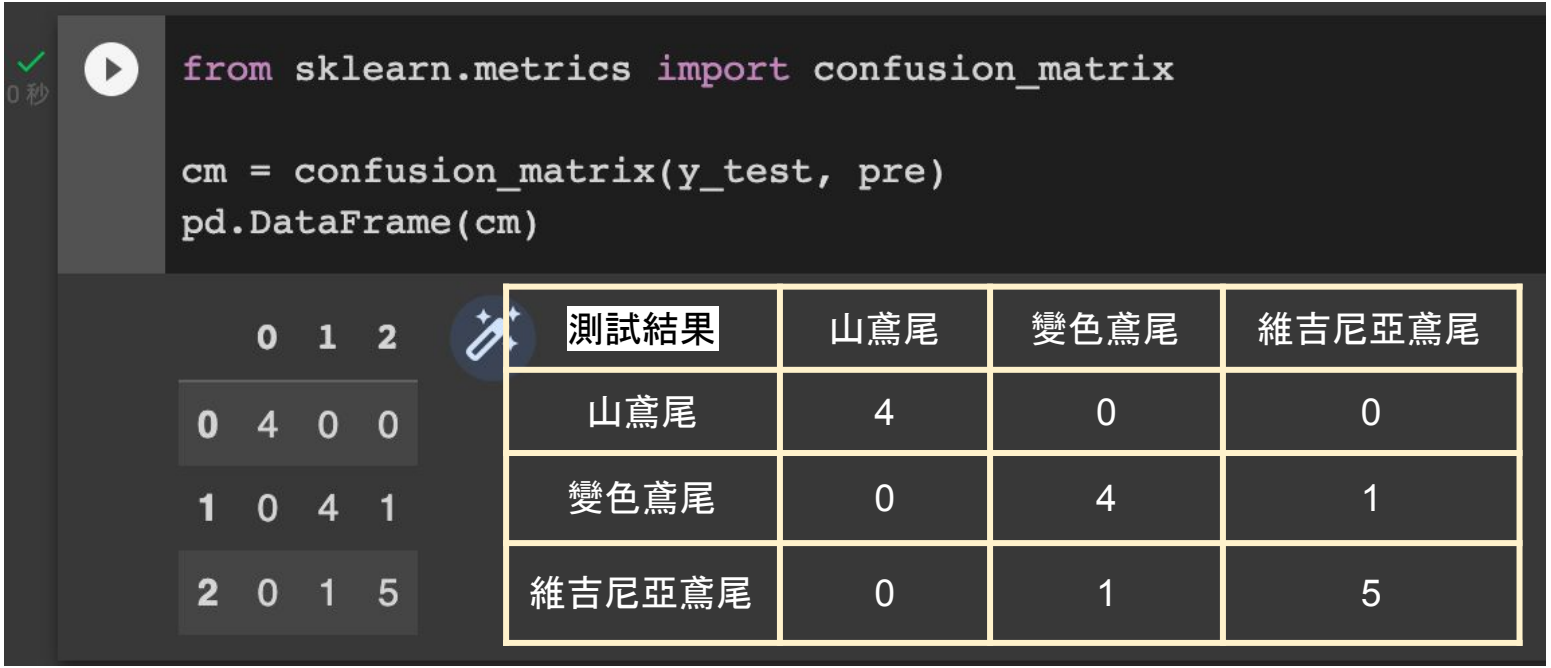
```
[146] from sklearn.metrics import accuracy_score

      rate = accuracy_score(pre, y_test)
      print("分類準確率:", rate)
```

```
分類準確率: 0.8666666666666667
```

多次測試後結果多落在 0.8~1.0 之間

衡量模型 - confusion_matrix



測試結果	山鳶尾	變色鳶尾	維吉尼亞鳶尾
山鳶尾	4	0	0
變色鳶尾	0	4	1
維吉尼亞鳶尾	0	1	5

DEMO

https://github.com/iamalex33329/iris_classification_tku

報告結束，謝謝大家