

- Ex1-12Add1:因為 iris 是知名的人門資料庫,是否有簡單方法就可印出 iris 的官方版介紹?如果有,請寫出這段程式。
 - Ex1-12Add2: Seaborn 是基于 matplotlib 的 Python 可視化庫。提供高級界面來繪制有吸引力的統計圖形。他的簡稱比較奇怪!在官方網站上建議的縮寫是什麼?以 iris 為例,如何快速畫出散點圖矩陣(pairplot)?在圖上,可見到任兩屬性的關係,至少可看出三種關係是「正相關」「無關」還有是什麼關係?
 - Ex1-12Add3:簡述三種 3 種容易取得的 Iris 資料集的來源,與相對應程式。
 -
-

Ex1-12Add1:因為 iris 是知名的人門資料庫,是否有簡單方法就可印出 iris 的官方版介紹?如果有,請寫出這段程式。

Ans:

有很多方法,例如:只要下列三行 python 即可:

```
from sklearn import datasets
iris = datasets.load_iris()
print(iris['DESCR'])
```

Ex1-12Add2: Seaborn 是基于 matplotlib 的 Python 可視化庫。提供高級界面來繪制有吸引力的統計圖形。他的簡稱比較奇怪!在官方網站上建議的縮寫是什麼?以 iris 為例,如何快速畫出散點圖矩陣(PairsPlots)?在圖上,可見到任兩屬性的關係,至少可看出三種關係是「正相關」「無關」還有是什麼關係?

Ans:

(1)正式的縮寫是 sns。

```
import seaborn as sns
```

(2)

在引入 seaborn 後,只要四行程式碼,就有可視化了!

以 iris 的散點圖矩陣(Pairs Plots)為例:

```
import seaborn as sns
sns.set(style="ticks")
df = sns.load_dataset("iris")
sns.pairplot(df, hue="species")
```

(3)

到任兩屬性的關係,至少可看出三種關係是「正相關」「無關」還有就是「負相關」。

Ex1-12Add3::簡述三種可取得 iris 資料集的方法,與相對應程式。

Ans:

(a)

方法一:用 sklearn (簡稱是 sk)

方法二:用 seaborn (簡稱是 sns)

方法三:到 iris 的官方網站 uci 大學

(b)

常用拿到 iris 的方法一:用 sklearn (簡稱是 sk)

```
from sklearn import datasets
import pandas as pd
import numpy as np
```

```
iris = datasets.load_iris()
print(iris['DESCR'])
===== 1=====
# 常用拿到 iris 的方法一：用 sklearn （簡稱是 sk）
from sklearn import datasets
import pandas as pd
import numpy as np
iris = datasets.load_iris()
#print(iris['DESCR'])
#print(iris['data'])
df = pd.DataFrame(data=iris['data'])
df.head()
x=iris['data']
y=iris['target']
merged_array=np.c_[x,y] # np.c_是更方便(不必調整成相同維度)就可 concatenate 實現陣列的水平
連接。
merged_array2 = np.concatenate((x, np.expand_dims(y, axis=1)), axis=1)
merged_array
df = pd.DataFrame(data=merged_array, columns=iris['feature_names'] + ['target'])
df.head()
===== 2=====
# 常用拿到 iris 的方法二：用 seaborn （簡稱是 sns）
import seaborn as sns
df=sns.load_dataset("iris")
#df.describe()
#df.info()
df[:3]
===== 3=====
# 到 iris 的官方網站 uci 大學
# QA 可查看第35個樣本的內容是 [[4.9 3.1 1.5 0.1 'Iris-setosa']] # 同理，那麼36,37,第38個樣本是？
import pandas as pd
df = pd.read_csv('https://archive.ics.uci.edu/ml/'
'machine-learning-databases/iris/iris.data', header=None)
#df.tail()
print("統計表(最大、最小、平均、樣本標準差)")
import numpy as np
df[:4]
```
