

# Day 58 非監督式機器學習

## 階層分群法 觀察： 使用 2D 樣版資料集

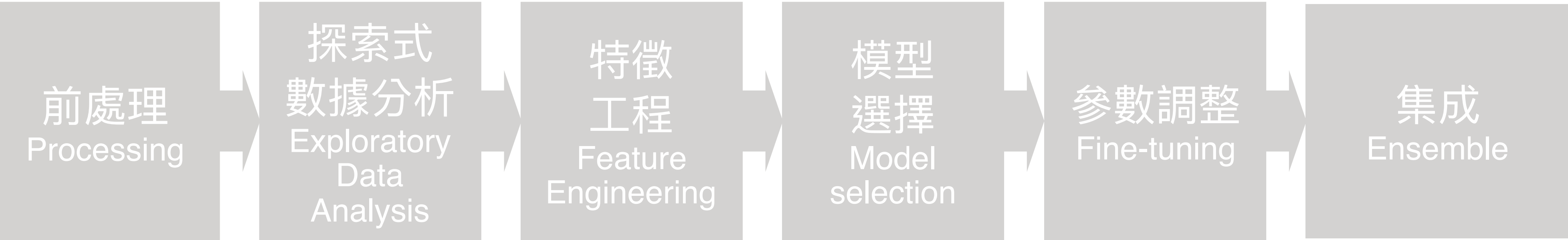


出題教練

周俊川 / 陳明佑

## 機器學習概論 Introduction of Machine Learning

監督式學習  
Supervised Learning



非監督式學習  
Unsupervised Learning



非監督學習  
Unsupervised learning

非監督簡介

分群 Clustering	K-平均算法 K-Mean
	階層分群法 Hierarchical Clustering
降維 Dimension Deduction	主成分分析PCA(Principal components analysis)
	T 分佈隨機近鄰嵌入 t-SNE



# 本日知識點目標

- 了解2D樣版資料集的設計原則與用途
- 大致知道有哪些套件有2D樣板資料集

註

因為非監督模型的效果，較難以簡單的範例看出來，所以非監督偶數日提供的檢視工具，僅供觀察非監督模型的效果，與後續其他部分及程式寫作無關，同學只要能感受到這些非監督模型的效果即可，不用執著於完全搞懂該章節所使用的工具

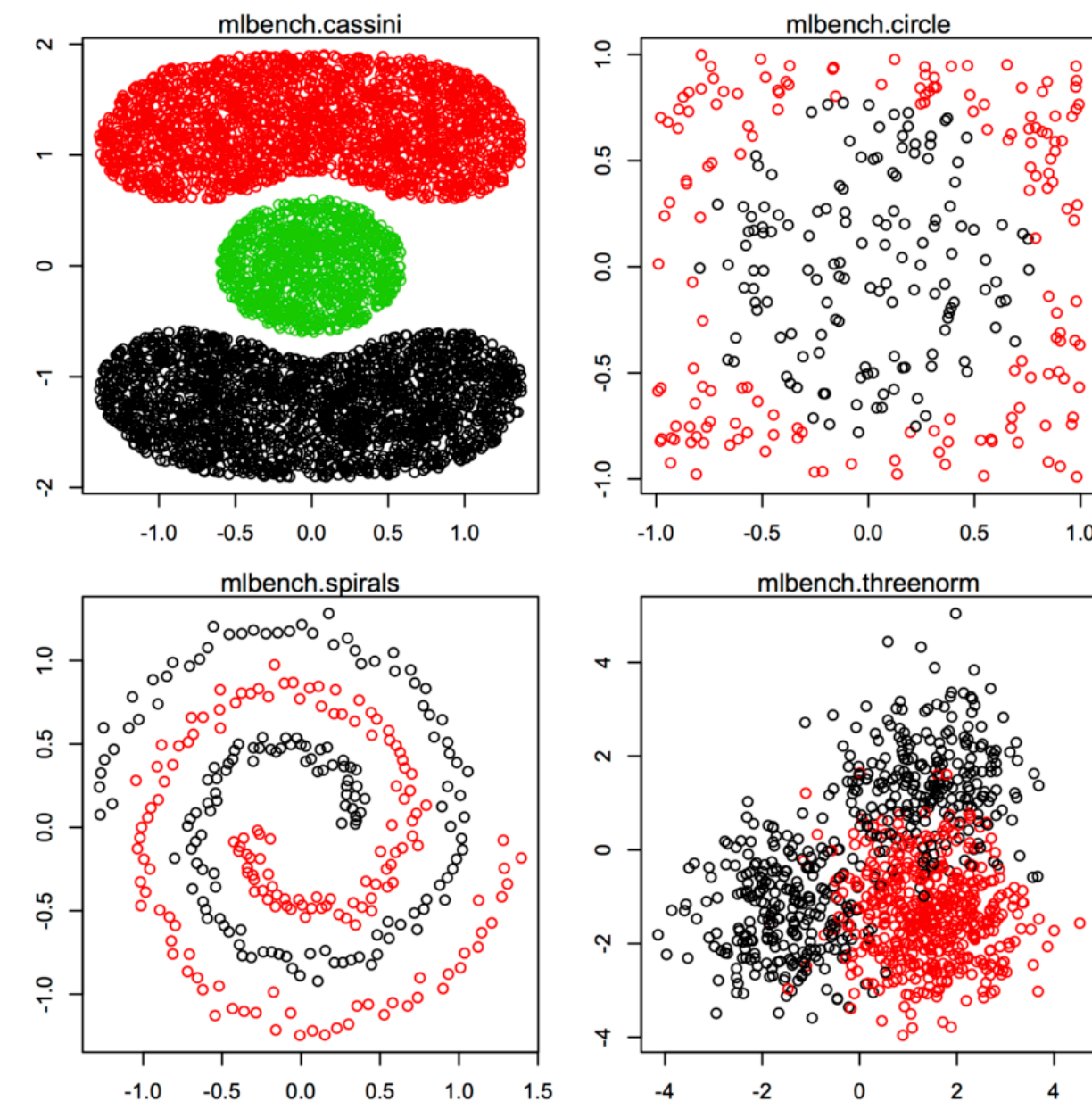
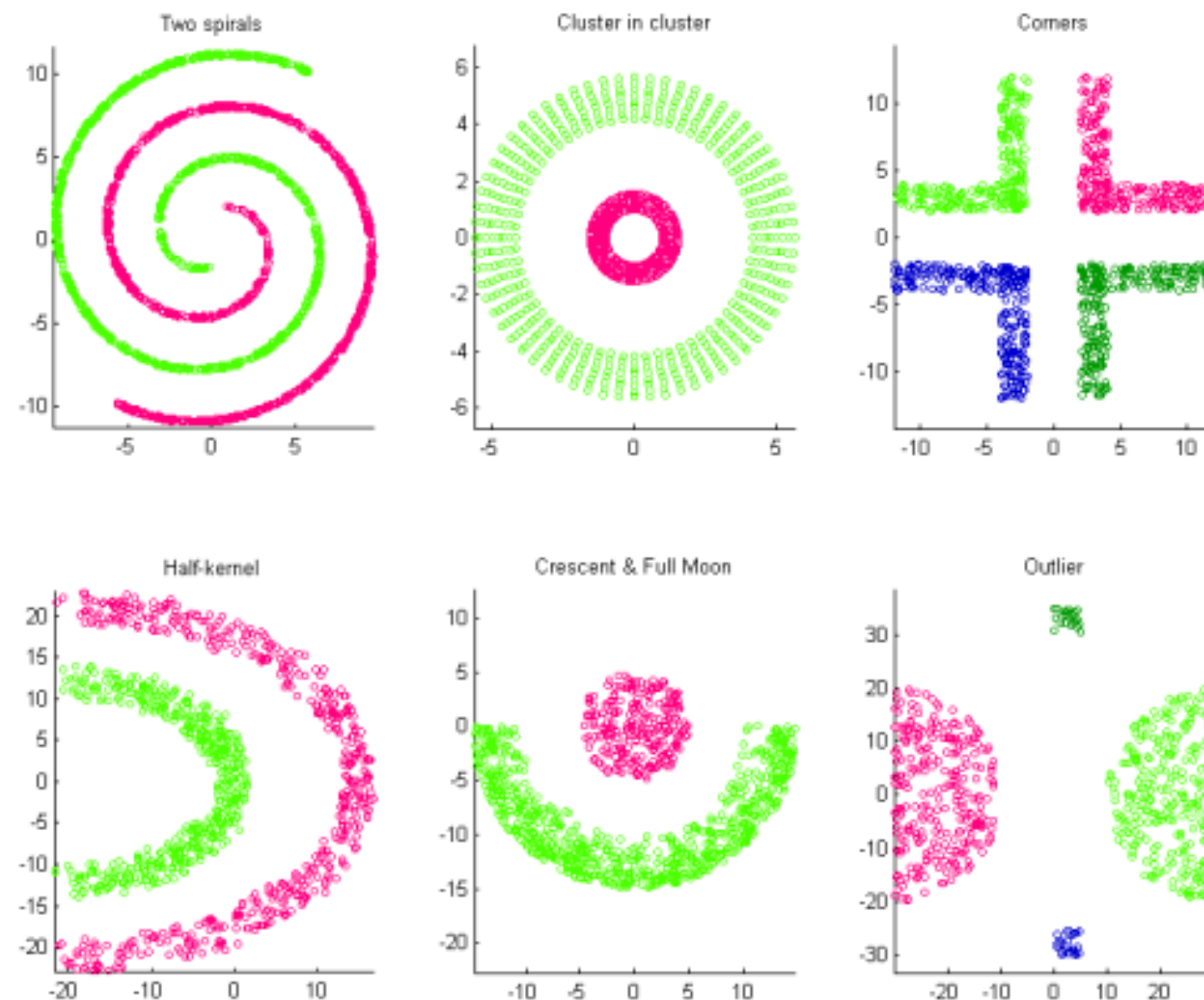
# 2D 樣版資料集 (2d toy dataset) ( 1 / 2 )

- 設計方式

與輪廓分析著重於以公式來衡量非監督模型不同，2D 樣版資料集著重於圖形的人機差異：挑選人眼容易分群，但非監督模型常常有困難的圖案樣板來展示

註

2d toy data / dataset 只是一個泛稱，並不只限於 sklearn 的 2D 樣板，例如下圖就是 mathworks 與 mlbench 的 2D 樣版資料

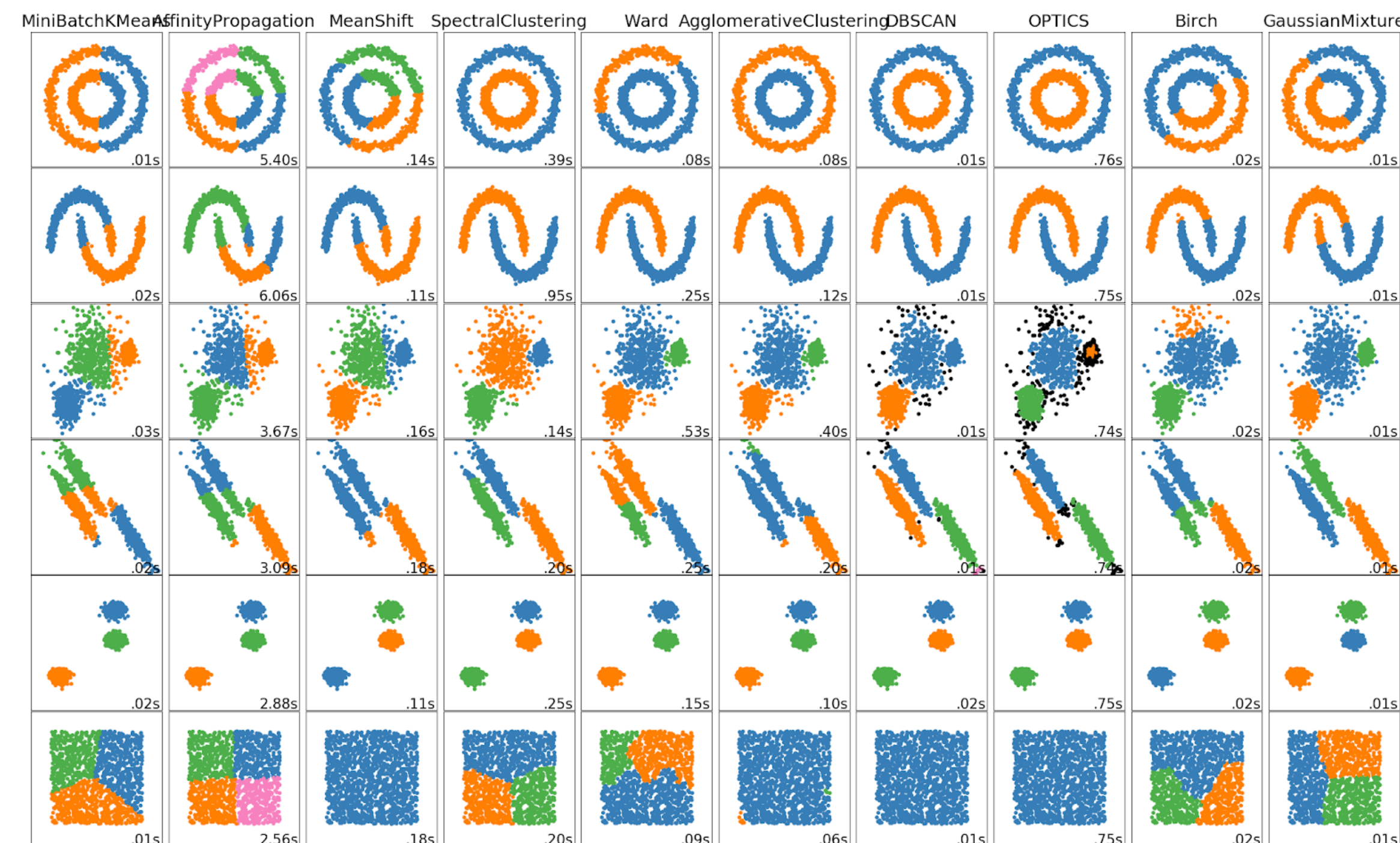




# 2D 樣版資料集 (2d toy dataset) ( 2 / 2 )

- 資料集用途

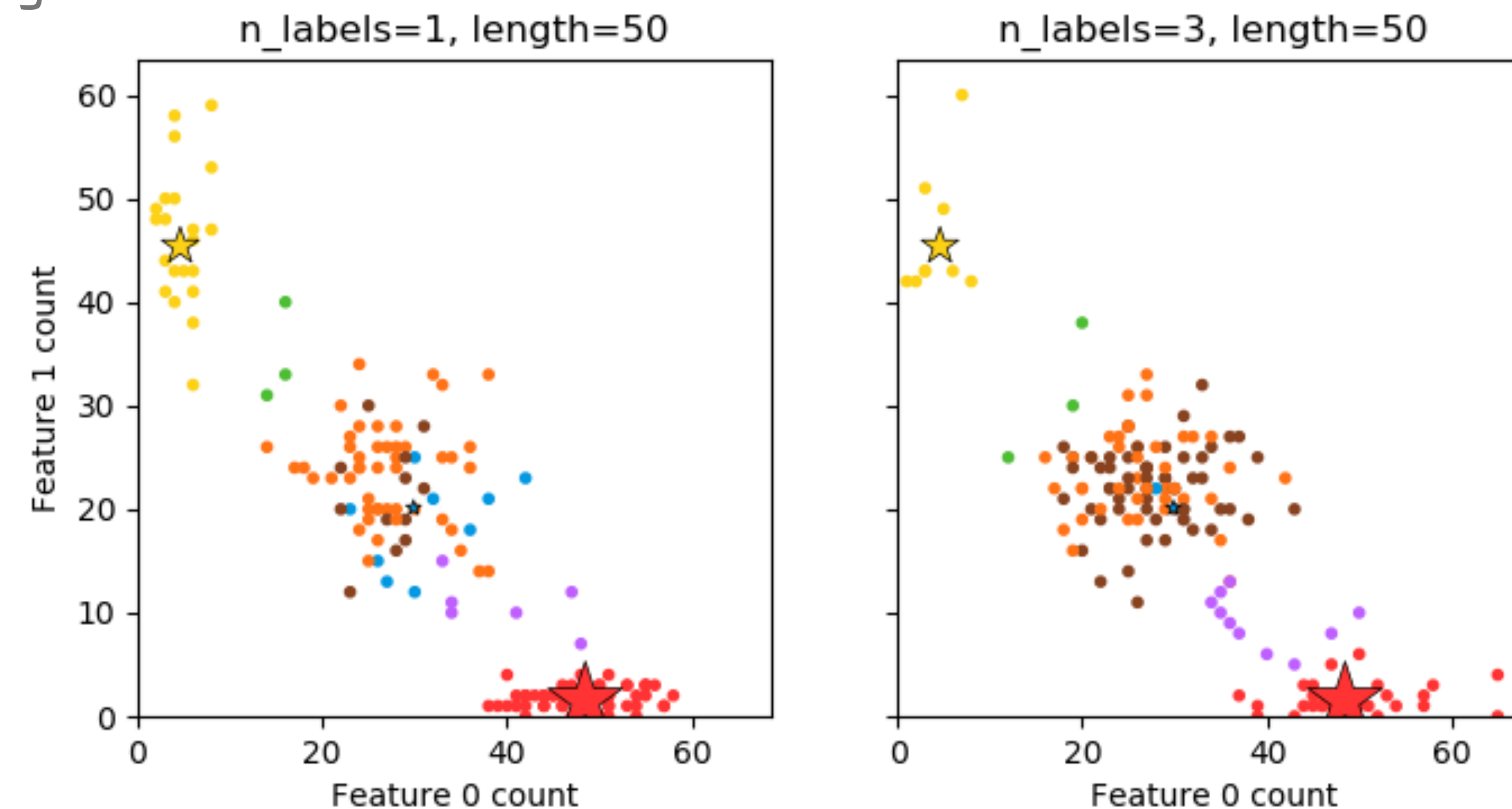
通常這樣的資料集，是用來讓人眼評估非監督模型的好壞，因為非監督模型的任務包含分群(對應於監督的分類)與流形還原(對應監督的迴歸)，所以2D 樣板資料集在設計上也包含這兩種類型的資料集

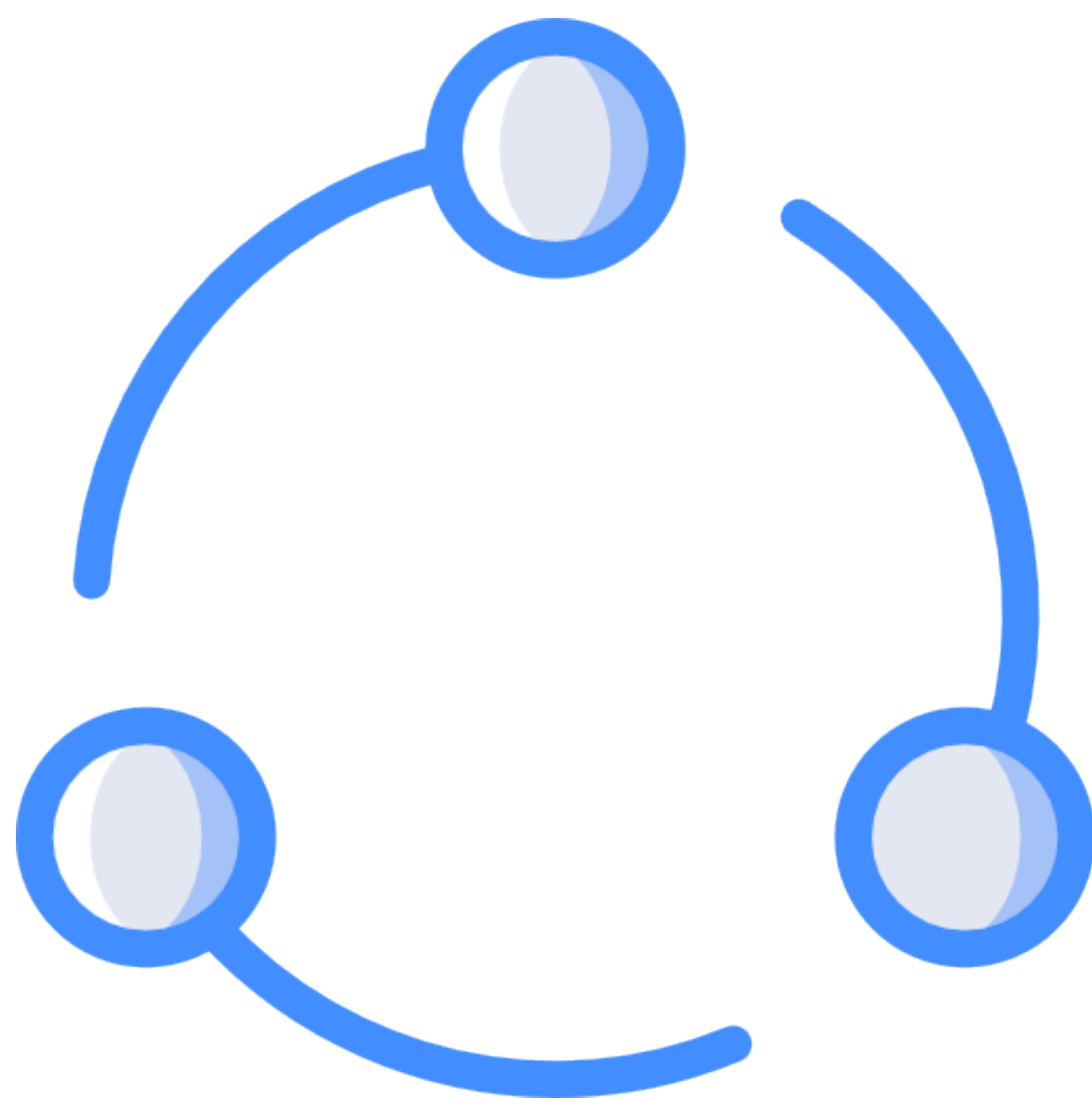


# sklearn 的 2D 樣版資料集

- sklearn 的 2D 樣版資料集

- sklearn 的資料集主要分為兩種：載入式 (Loaders) 與生成式 (Samples generator)，載入式的是固定資料，生成式的則是先有既定模式，在模式下有限度的隨機生成每次使用的資料集
- 2D 樣版資料集屬於生成式資料集(如右圖)，使用不同分布，用以顯示各種非監督模型的優缺點，提供使用者參考





- 了解2D樣版資料集的設計著重於圖形的人機差異，用途在於讓人眼以非量化的方式評估非監督模型的好壞，也因為非監督問題的類型不同，這類資料集也有分群與流形還原等不同對應類形
- 2D樣板資料集很多套件都有，也不限於只有Python 上使用的套件：如 sklearn / mathworks / mlbench 都有對應的資料集



# 解題時間 Coding Time

請跳出PDF至官網Sample Code & 作業  
開始解題

