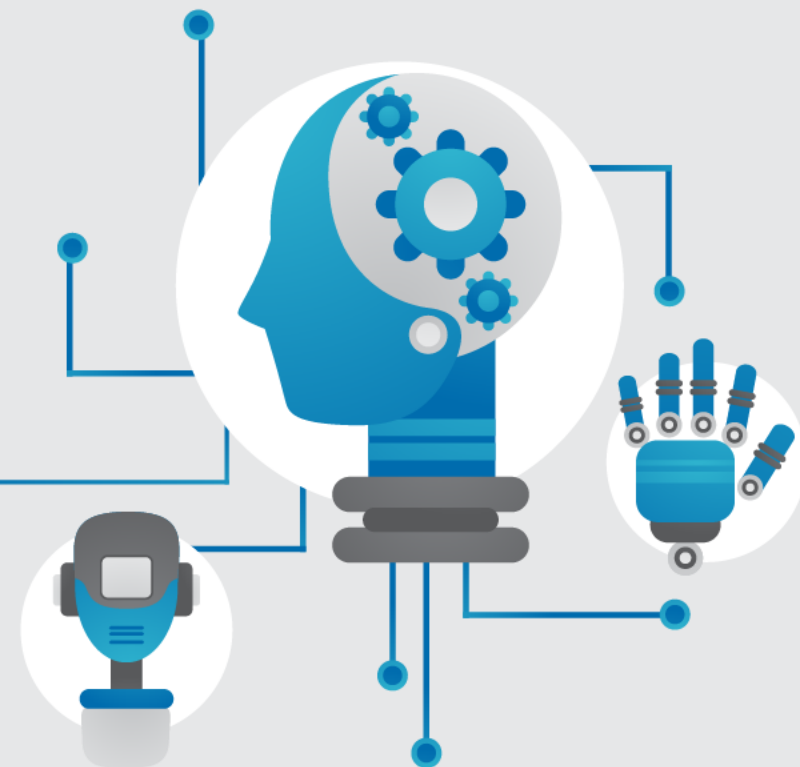
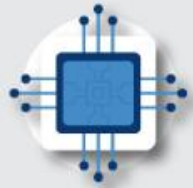


# DBSCAN 分群法





# DBSCAN 演算法



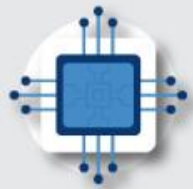
- › DBSCAN ( Density-Based Spatial Clustering of Applications with Noise ) 演算法是在 1996 年由Martin Ester, Hans-Peter Kriegel, Jörg Sander 及 Xiaowei Xu 提出的分群演算法。
- › DBSCAN演算法的分群依據為**密度**，不需要設定群組數，因此相較於K-Means能更好的找出不同形狀的資料群，且能找出**離群點**（最接近它的點也很遠）。



# DBSCAN 演算法



- › DBSCAN演算法有**偵測距離**與**最少樣本數**兩個參數。
- › 以某資料點為中心，在其周圍的偵測距離範圍內若有最少樣本數以上個資料點，則此區域稱為**密集區**（Dense region）。該資料點稱為**核心樣本**或**核心點**（Core samples, core points）。
- › 位於某資料點為中心的密集區但不符合核心樣本的資料點稱為**邊界點**（Boundary points）。
- › 不在任何密集區的資料點被歸類為**雜訊**（Noise），亦即不屬於任何群組。



# DBSCAN 演算法步驟



## › 演算法步驟

1

挑選一個無標記的資料點 $v$ ，如果是核心點，則給一個新的群組標記並進行下個步驟，否則標記為雜訊並重複此步驟。

2

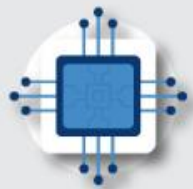
拜訪此資料點的所有鄰居（距離小於偵測距離 $r$ ），若尚無標記，則設定與 $v$ 相同標記。

3

如果鄰居資料點也是核心點，以該鄰居資料點為中心重複步驟2，直到沒有其它核心點。

4

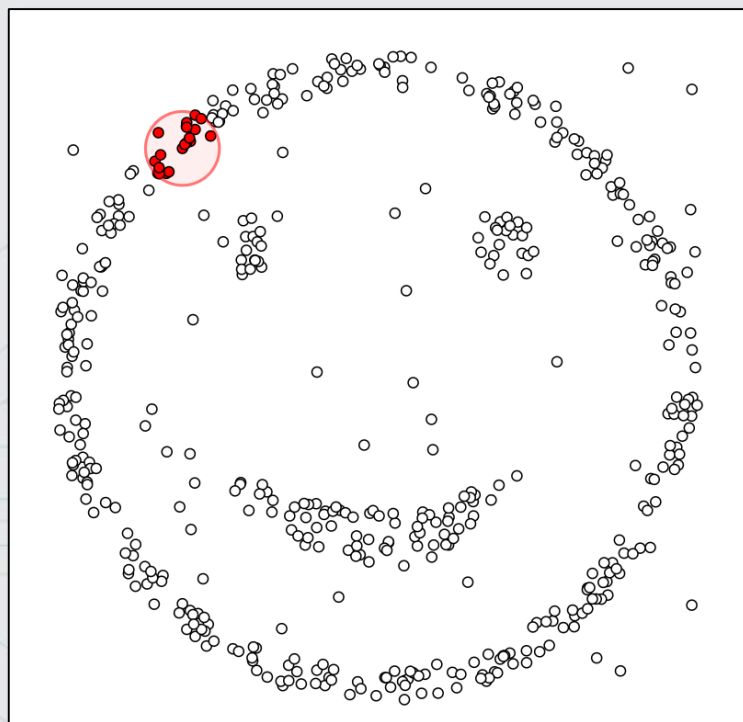
重複步驟1~3，直到所有資料點都處理完畢。



# DBSCAN 範例說明



- › 從資料點中選一個未標記的點
  - 如果是核心點，則把範圍內資料點標記為同一群組。
  - 如果不是核心點，則設為雜訊，繼續搜尋。

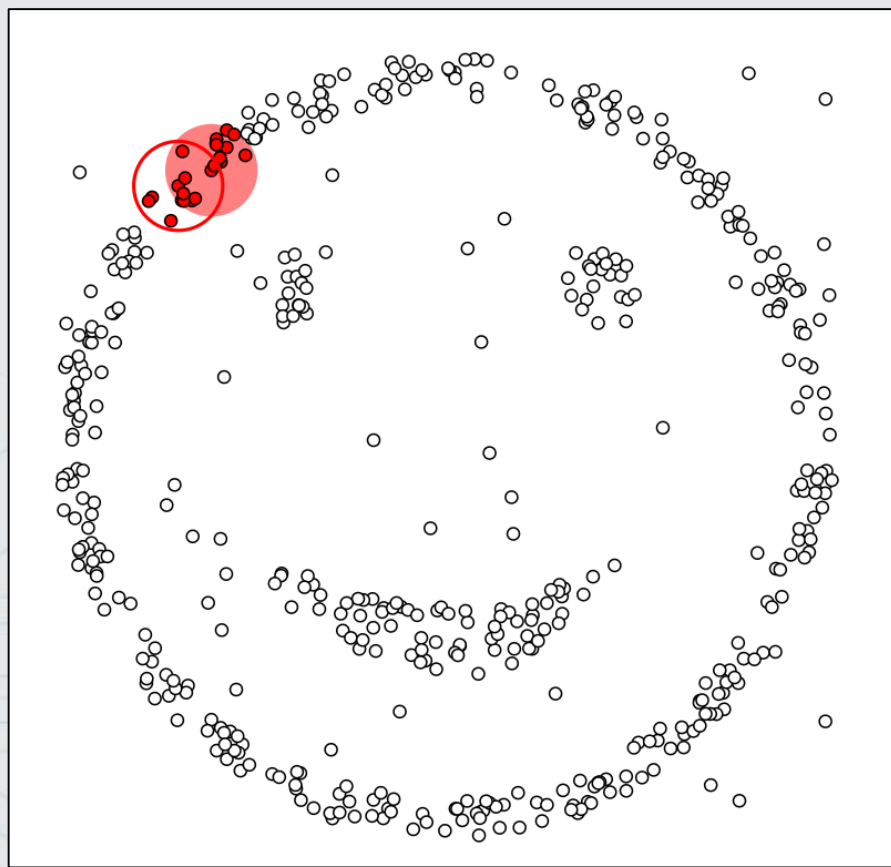


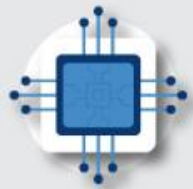


# DBSCAN 範例說明



- › 檢查所有鄰居，如果鄰居也是核心點，則重複直到沒有鄰居為核心點

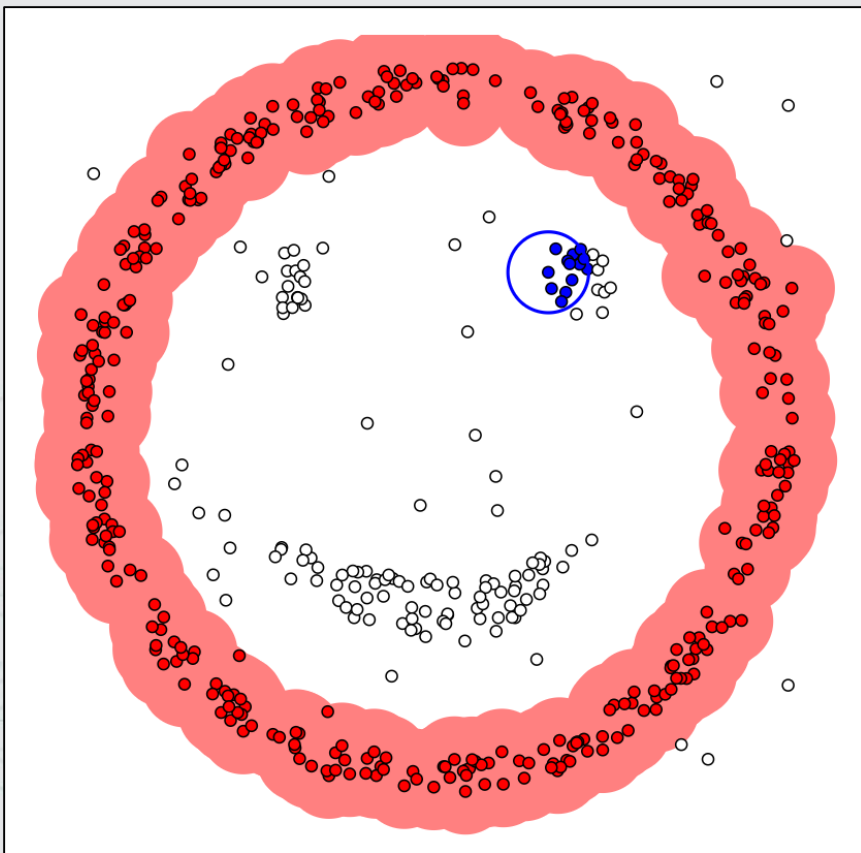




# DBSCAN 範例說明



› 重複步驟1 ~ 3，找到下一個未標記點，重新開始





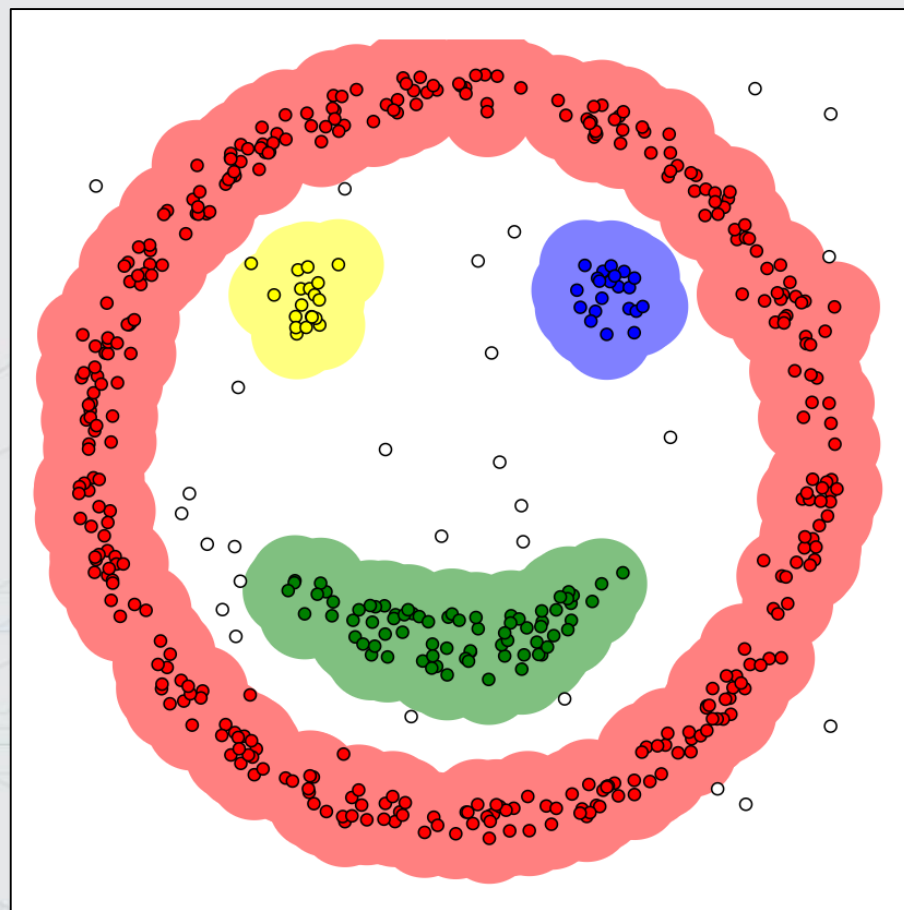


# DBSCAN 範例說明

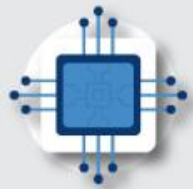
機器學習實務



› 直到掃描完所有點







# 優缺點



## 優

- 不用設定群組數
- 能找出離群點（噪點）
- 能適用比較複雜的資料形狀

## 缺

- 參數設定不好會影響效果甚大
- 如果資料點有不同的密度，且差異很大，DBSCAN 將無法提供一個好的分群結果