

16-3 K-means演算法

16-3-1 認識K-means演算法

16-3-2 使用K-means演算法分群鳶尾花

16-3-1 認識K-means演算法 – 說明

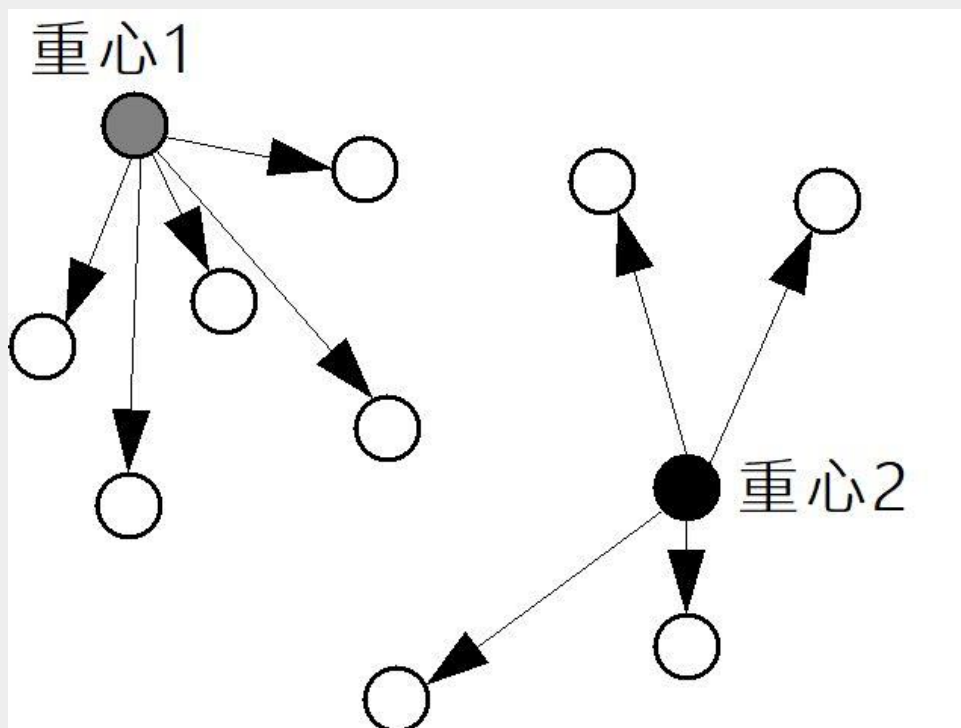
- **分群和分類的差異**在於：分類是在已知資料集分類的情況下，替新東西進行分類，分群是在根本不知資料集分類的情況下，直接使用特徵來進行分類，K-means就是機器學習常用的一種分群演算法。
- **K-means分群**（ K-means Clustering ）也稱為K平均數分群，因為我們並不用知道資料集分類的情況下，即可進行分群，這是一種非監督式學習（ Unsupervised Learning ）。

16-3-1 認識K-means演算法 – 基本步驟(Step 1)

- K-means分群的作法是先找出K個群組的重心 (Centroid)，資料集就以距離最近重心來分成群組後，重新計算群組的新重心後，再分群一次，重複操作來完成分群，其步驟如下所示：
- Step 1：依資料集數決定適當的K個重心，例如：2個。

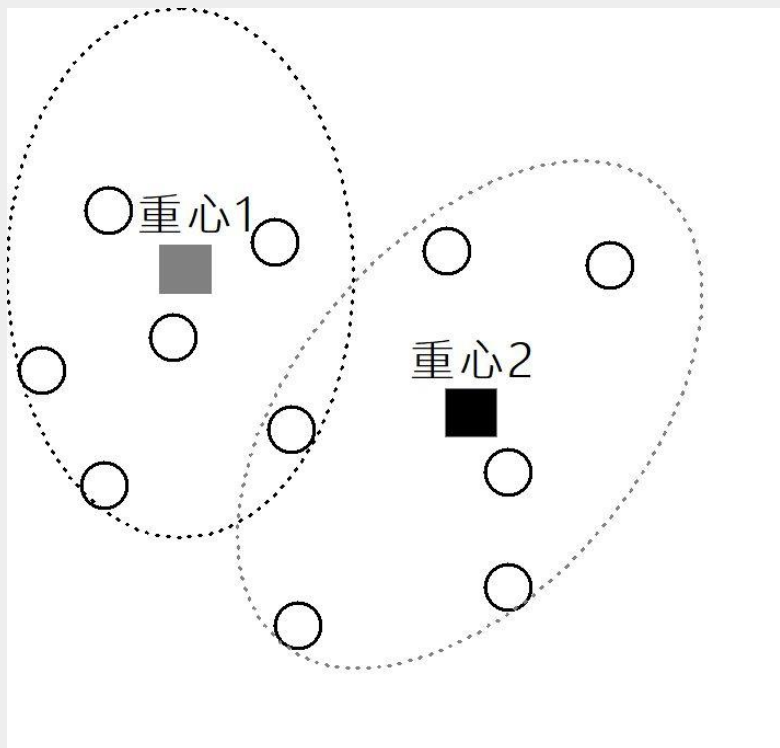
16-3-1 認識K-means演算法 – 基本步驟(Step 2)

- Step 2：計算資料集和重心的距離（公式和K鄰近演算法相同），然後以距離最近重心的資料來分成群組，如下圖所示：



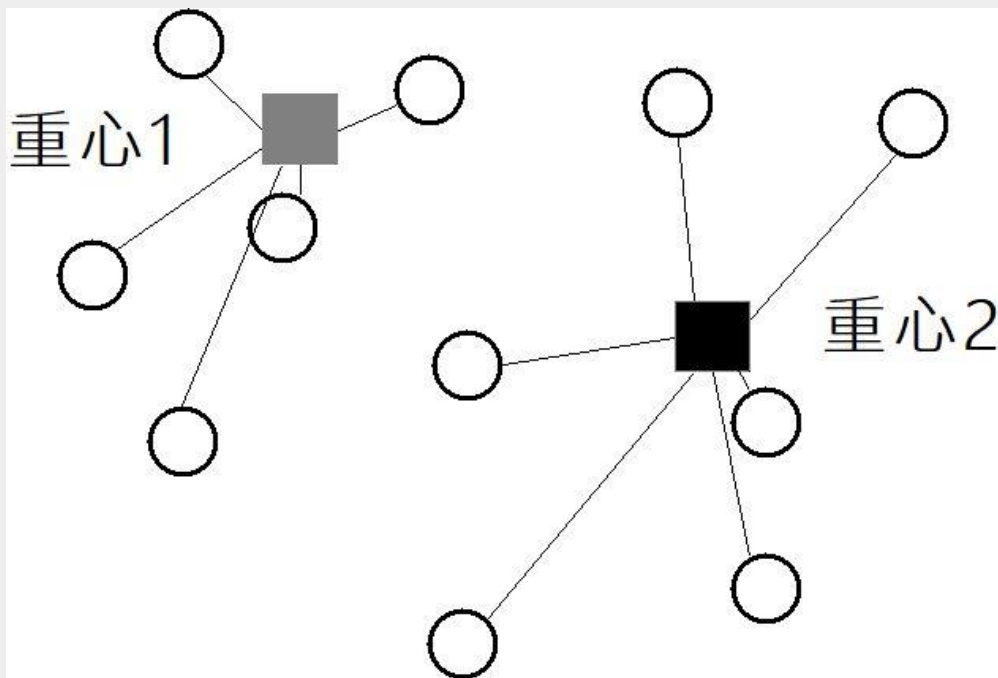
16-3-1 認識K-means演算法 – 基本步驟(Step 3)

- Step 3：重新計算群組資料集各特徵的算術平均數作為新的重心，如下圖所示：



16-3-1 認識K-means演算法 – 基本步驟(Step 4)

- Step 4：再次計算資料集和重心的距離，然後以距離最近重心來分成群組，如下圖所示：



16-3-1 認識K-means演算法 – 基本步驟(Step 5)

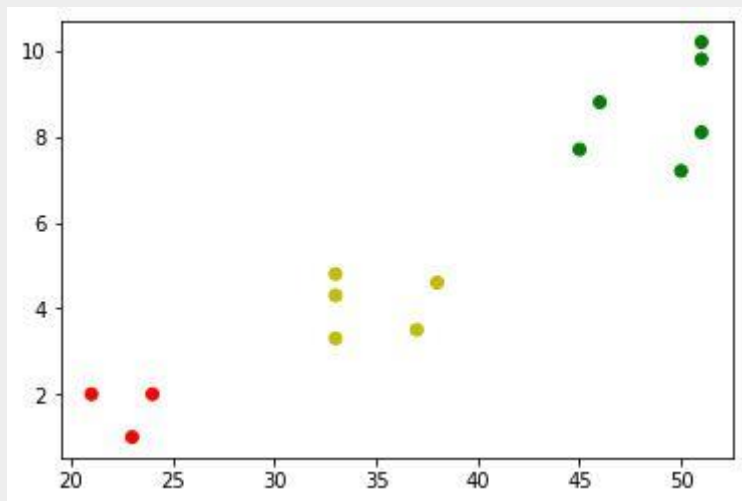
- Step 5：重複操作Step 3~4直到重心和群組不再改變為止。

16-3-1 認識K-means演算法 – 範例：依據動物的體重和身長來分群

- 在動物園收集到14隻動物的體重和身長資料，如下表所示：

身長	51	46	51	45	51	50	33	38	37	33	33	21	23	24
體重	10.2	8.8	8.1	7.7	9.8	7.2	4.8	4.6	3.5	3.3	4.3	2.0	1.0	2.0

- 在K值3的情況下，請使用K-means演算法替14隻動物進行分群，如下所示：



16-3-2 使用K-means演算法分群鳶尾花

- 在第16-2-2節是使用K鄰近演算法分類鳶尾花，和使用散佈圖來視覺化顯示鳶尾花資料集，這一節我們準備改用K-means演算法來分群鳶尾花，事實上，這也是在分類鳶尾花。
 - 建立K-means模型分群鳶尾花
 - 修正分群標籤錯誤重繪散佈圖
 - K-means模型的積效測量