

一、Introduction

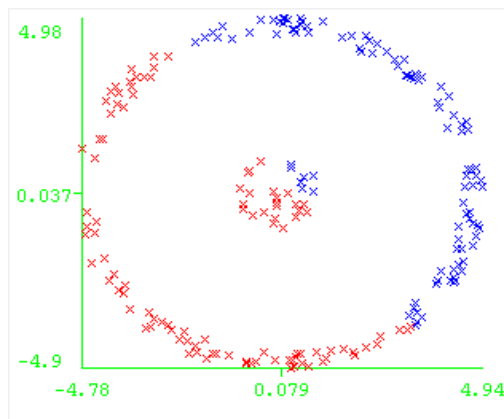
1. 使用 Data set : Concentric circles
2. 資料維度：205 * 2 (205 在 X-Y 平面上的點,資料均為數字)
3. 資料視覺化：兩個同心圓
4. 實驗目的：使用分群演算法，並透過調整 K 值和距離計算方式，嘗試將兩同心圓分成兩群或使小圓與大圓分屬不同群。

二、Methods

使用演算法： K-Means , Hierarchical Clustering

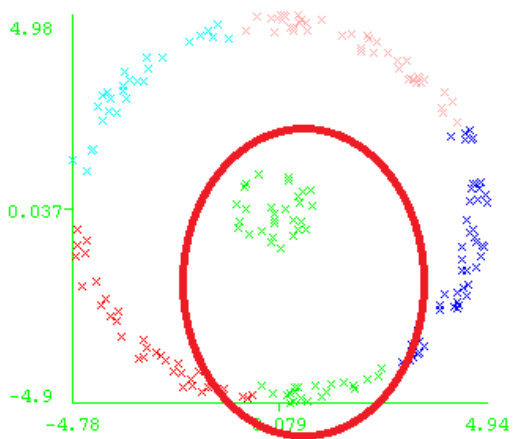
三、Experiments

1. K-Means (K=? , 歐氏距離 / 曼哈頓距離) :



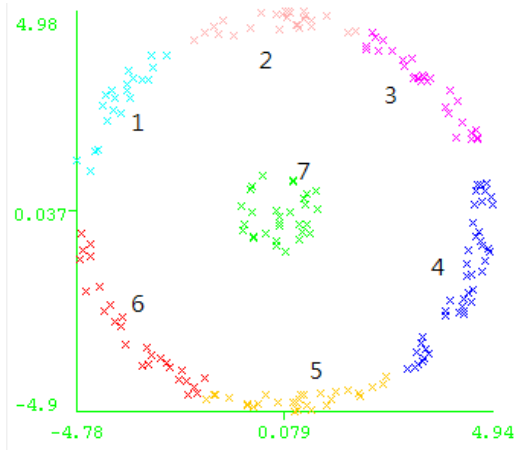
← **K=2, 歐氏距離**

分群失敗，中間與外圍分成同一群。
正確率 = $(205 - 30) / 205 = 85.3\%$
(判定中間部分均為分錯的資料)



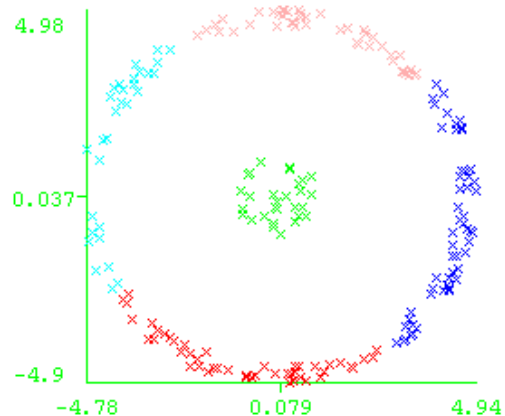
K=5, 歐氏距離

分群失敗，中間與外圍分成同一群。
正確率 = $(205 - 21) / 205 = 89.7\%$
(判定下方綠色為分錯的資料)



← **K=7, 歐氏距離**

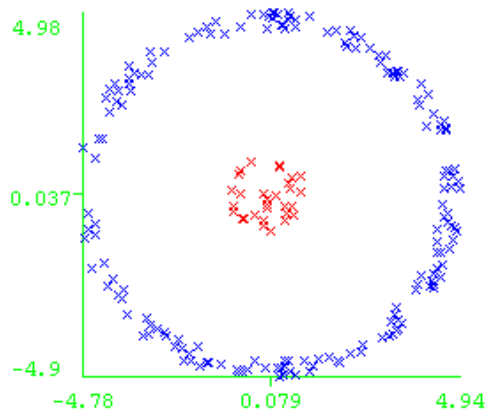
分群成功，中間部分不與外圍部分重疊。
正確率 = 100%



K=5, 曼哈頓距離

分群成功，中間部分不與外圍重疊。
正確率 100%

2. Hierarchical Clustering (cluster = ? , single / average linkage) :

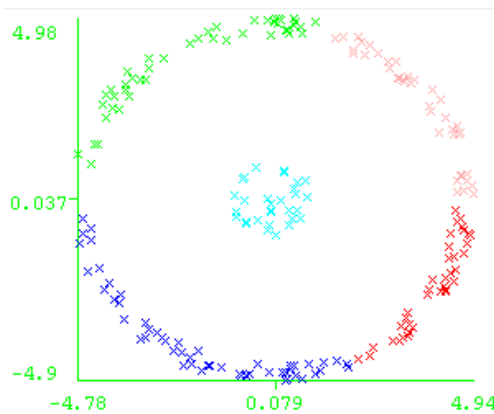
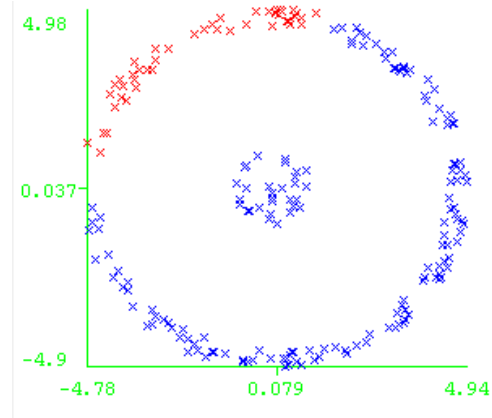


← **cluster = 2 , single linkage**

已可以分群成功。

正確率 100%

cluster = 2 , average linkage →
分群失敗，中間與外圍分為一群
正確率 = $(205 - 75) / 205 = 63.4\%$
(中心與外圍紅色區域視為錯誤)



← **cluster = 5 , average linkage**

已可以分群成功。

正確率 100%

四、Discussion

1. K-means :

(K=5 , 曼哈頓距離) , (K=7 , 歐氏距離)

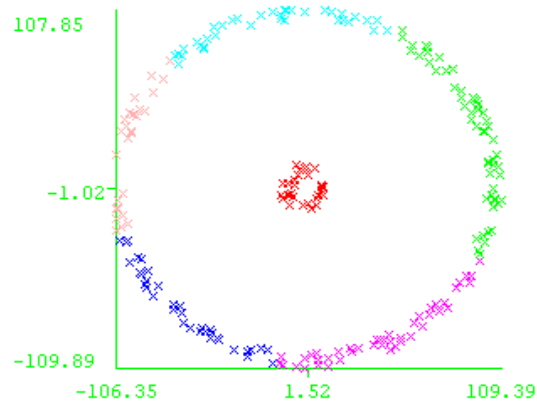
Hierarchical Clustering :

(cluster=2 , single linkage) , (cluster=5 , average linkage)

以上四種組合皆可將資料分為內圈與外圈，不會有內圈與外圈分為同一群的情況。

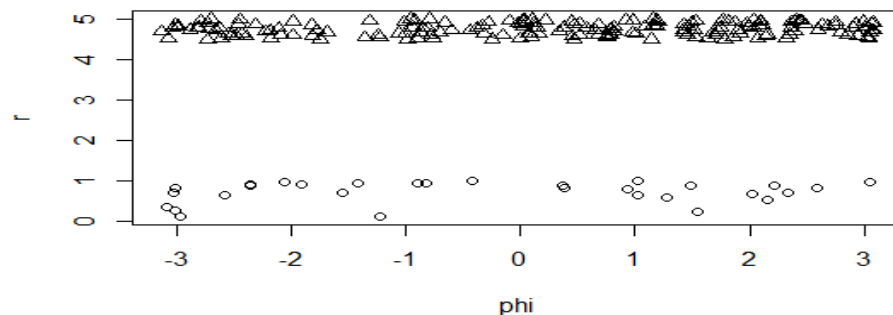
- 經由以上的實驗過程，從 2 開始嘗試，分群失敗則增加 K 值，其他參數不變，直到找到能正確分群的 K 值為止，而由於分群的結果與對距離的定義有很大的關聯，若是兩同心圓的間距變小的時候，勢必會讓 K 值增加，來降低每個群的半徑，讓中間的圓不會被外圍的群分在一起，反之若同心圓間距變大，則 K 值將會降低，外圍群的半徑將會變大。

3. 所有實驗過程中，最理想的結果是 Hierarchical Clustering (cluster=2 , single linkage)這個組合，可以完美的將同心圓分為兩群(內圓與外圓)，原因為 Hierarchical 會將所有資料中最相鄰的兩點(群)變為一群，直到剩下 2 群為止，又因為 single linkage 找兩群之間最相近的點，而此時內圓與外圓中間有間距，且內圓的點相對密集，會先被分為一群，這樣的條件之下顯然內外圓就不可能合併為一群。
4. 將兩圓的間距加大 (內圓半徑約為 5，外圓半徑約為 100)，間距加大使用 Kmeans 只需 6 群即可分成內外圓 (原本要 7 群)。



五、利用 R 轉換 Domain

1. 使用 Rstudio , library : pracma 內的 cart2pol()函數轉換為向量表示 (phi , r) 由於內圓半徑較小，外圓半徑較大，視覺化之後如下圖：



下半部為內圓的資料，上半部為外圓資料，利用 Kmeans (K=2)，即可正確分為兩群。

六、References

1. hierarchical clustering (Wikipedia)
https://en.wikipedia.org/wiki/Hierarchical_clustering
2. k-means clustering
https://en.wikipedia.org/wiki/K-means_clustering
3. R – as.numeric
<http://stackoverflow.com/questions/28962152/shapiro-test-is-numeric-x-is-not-true>