

環境與程式語言

RStudio, R 3.1.3, Windows 7

R dependent library

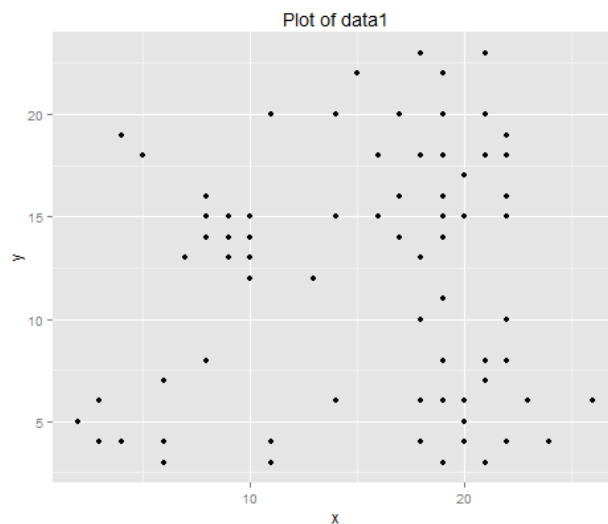
```
RWeka // install with following command > install.packages("RWeka")  
ggplot2 // plot library
```

使用 kmeans 作為分群的演算法，要決定分成幾群，可以先將 test data 畫圖，再決定分成 n 群，將結果存成 txt 檔，步驟：

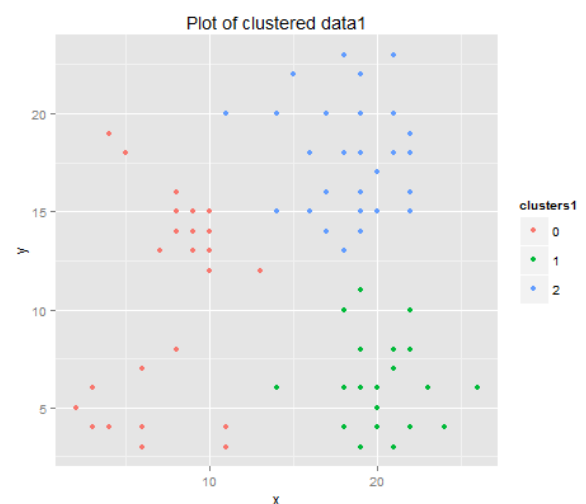
1. Plot training data set
2. 決定分成 n 群 (n 視 training data set 分布情形而定)
3. 開始使用演算法分群
4. 輸出 output.txt 檔

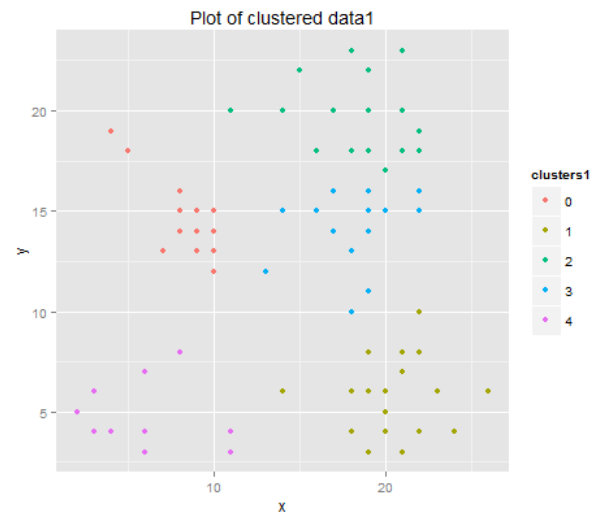
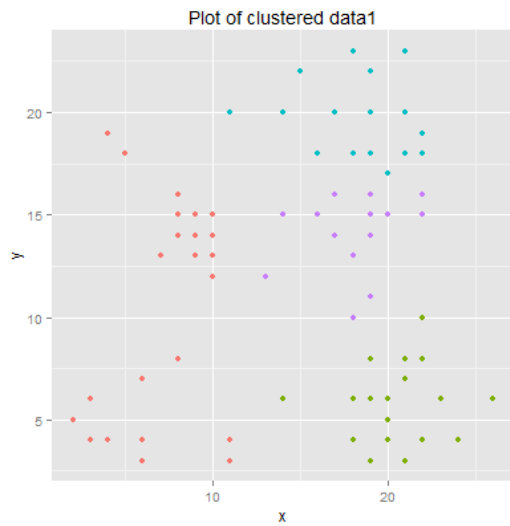
■ Data set 1 - 0_random.txt

1.



2. 決定分別分成 2, 3, 4, 5 群
- 3.



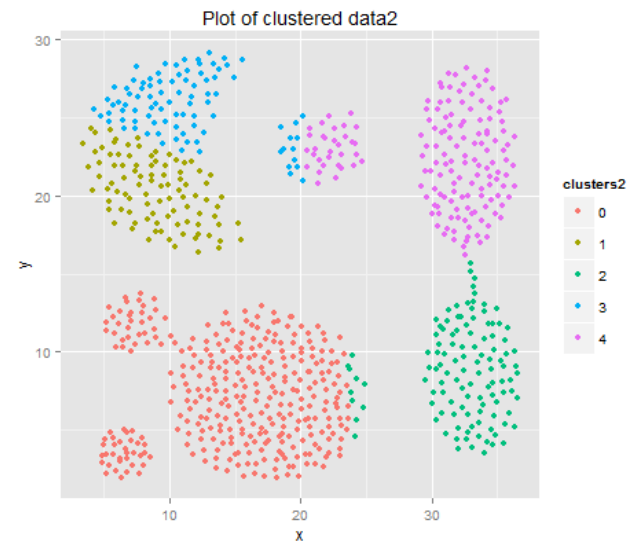


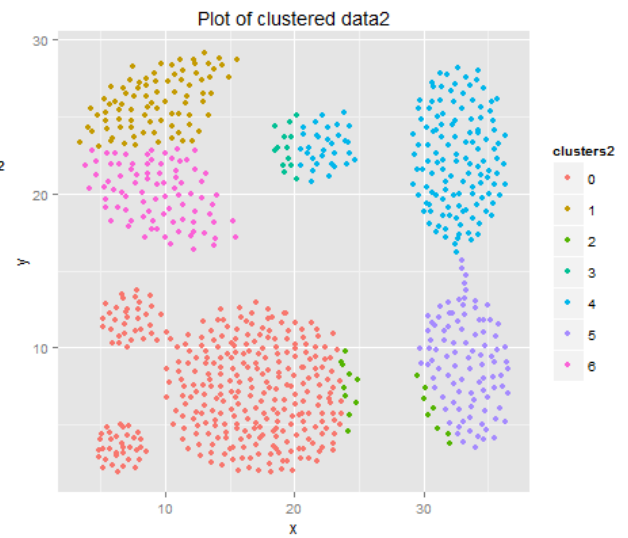
■ Data set 2 - 1_random.txt

- 1.



2. 決定分別分成 4,5,6,7 群
- 3.





使用 kmeans 來對 data set 2 作分群結果不太理想，可以考慮使用 density-based cluster algorithm。

■ Output

最後 data set1 and data set2 輸出的檔案為 output1.txt 及 output2.txt。

■ Execute

```
> source("hw4.R")
```