

日期: 2018/12/19(三) 上機考: Open Book
授課教師: 吳漢銘 (臺北大學統計學系副教授)

請仔細閱讀每一個注意事項 (禁止討論)

1. 考試期間

- (a) 請按照平時上課之座位入座。
- (b) 可參考課本、上課講義 (包含電子檔) 及其它資料, 但不能與別人討論。
- (c) 可使用計算機、自己的筆記型電腦及平板電腦, 不可使用手機。
- (d) 全程可上網查詢, 但不能用通訊軟體 (例如: FB/LINE) 討論, 也不可抄襲網路上之程式碼。
- (e) 有問題者, 請舉手發問。勿與同學交談。
- (f) 不按照規定作答者, 酌量扣分。
- (g) 不可使用它人之隨身碟。「作弊」或「疑似作弊」, 往後各項考試不予評分。

2. 下載題目卷, 上傳答題檔案:

- (a) 於課程網站下載題目卷。
- (b) 上傳答題檔案: 於教師網站首頁登入 [作業考試上傳區], 帳號: hdda。密碼: xxx (上課教室號碼)。
- (c) 請上傳「學號-姓名-HDDA-exam2.docx」。(目錄: 「20181219-exam2」。)

3. 答題檔案原則:

- (a) 請依照「R 程式作業繳交方式」, 複製 Console「程式執行及結果」至答案卷。圖形複製, 請注意大小, 內容數字文字需可辨識。
- (b) 程式設計題, 若程式碼直接複製 (或照抄) 講義上的以不給分為原則。
- (c) 若上傳檔案格式錯誤, 內容亂碼, 空檔等等問題。請自行負責。
- (d) 若要重覆上傳 (第 2 次以上), 請在檔名最後加「-2」、「-3」, 例如: 「學號-姓名-HDDA-exam2-2.docx」、「學號-姓名-HDDA-exam2-3.docx」等等。
- (e) 上傳兩次 (含) 以上、格式不合等等酌量扣分。

4. 完成考試

- (a) 上傳完畢, 請通知教師確認。
- (b) 確認無誤, 請刪除作答目錄及答案卷, 清空資源回收筒, 並關機。即可離席。

我已經仔細閱讀上述各注意事項, 若有違背, 會自行負責。

日期: 2018/12/19(三) 上機考: Open Book
授課教師: 吳漢銘 (臺北大學統計學系副教授)

1. 以下三個 R 套件皆提供一些函數可計算 Shrinkage estimation of covariance matrix:

- cov.shrink {corpcor}
- shrinkcovmat.identity {ShrinkCovMat}
- covEstimation {RiskPortfolios} with type = 'oneparm'

現以 R 程式產生一模擬資料 x 如下:

```
library(MASS)
n <- 10
p <- 100
set.seed(123456)
sigma <- matrix(rnorm(p * p), ncol = p)
sigma <- crossprod(sigma) + diag(rep(0.2, p))
x <- mvrnorm(n, mu=rep(0, p), Sigma=sigma)
```

試以不同的 p/n 值 ($p/n = 0.1, 0.5, 2, 10$, p 固定為 100) · 繪圖 (一頁 4 張圖) 比較不同 Shrinkage 方法所計算出來的共變異數矩陣之 eigenvalues, 同時也需與真實共變異數矩陣的 eigenvalues 及傳統共變異數矩陣的 eigenvalues 相比較 (參照 DR 講義 117-119/144)(一張圖上共 5 條線, 每一條線代表一個共變異數矩陣計算方法)。

2. 承上題之 4 組模擬資料 ($p/n = 0.1, 0.5, 2, 10, p = 100$)(各命名為 $x.pn0.1, x.pn0.5, x.pn2, x.pn10$)，以主成份分析 (PCA) 及等軸距特徵映射 (ISOMAP) 做維度縮減 (維度為 2~5)，計算 LCMC 評估指標 (其中鄰居個數 (neighborhood size) 範圍為 5~10)。(使用之方法參數，請自行選合適的。PCA/ISOMAP 請說明所使用的套件或修正方法)，將結果以下列格式列出：

x.pn0.1						
PCA						
dim	K=5	K=6	K=7	K=8	K=9	K=10
2
3
4
5
ISOMAP						
dim	K=5	K=6	K=7	K=8	K=9	K=10
2
3
4
5
x.pn0.5						
PCA						
dim	K=5	K=6	K=7	K=8	K=9	K=10
2
3
4
5
ISOMAP						
dim	K=5	K=6	K=7	K=8	K=9	K=10
2
3
4
5
...						

3. **資料集: Morphological Measurements on Leptograpsus Crabs:** The crabs data frame has 200 rows and 8 columns, describing 5 morphological measurements on 50 crabs each of two colour forms and both sexes, of the species *Leptograpsus variegatus* (紫岩蟹) collected at Fremantle, W. Australia.

```
library(MASS)
data(crabs)
names(crabs)
?crabs
crabs2 <- crabs
crabs2$index <- rep(1:4, each=50)
```

- (a) 以 pheatmap 套件畫出資料 crabs2 ("FL", "RW", "CL", "CW", "BD") 未分群前之熱圖 (heatmap)。(熱圖邊需加上"sp", "sex", "index" 之資訊，且需有圖例說明 (legend))
- (b) 對此資料 (200 隻蟹) 做 K 均值法群集分析 ($K = 4$)，5 個欄位變數則做 average-linkage，將分群結果，一同呈現在上述的 heatmap 中。(即新增一彩色邊條為 K 均值法之結果，整個資料熱圖之排序，依序是 K 均值法結果、"sp"、"sex"、"index"。)(資料排序提示: arrange {dplyr})
- (c) (承 (a) 小題) 以 pheatmap 套件畫出資料 crabs2 (蟹及變數) 之 complete-linkage 和 single-linkage 之熱圖。(熱圖邊需加上"sp", "sex", "index" 之資訊，且需有圖例說明 (legend))(需將兩熱圖畫在一起比較，即一頁兩張圖) (熱圖請使用與 (a) 小題不同之色階)

注意: 上傳檔案之後，請刪除作答目錄 及答案卷，清空資源回收筒，關機。交回題目卷。