

繳交截止日期: 2020/06/20(五), 24:00

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. 寫作業要點

- (a) 可參考課本、上課講義 (包含電子檔) 及其它資料。
- (b) 儘量不要與別人 (或同學) 討論, 自己做, 不可參考同學的答案, 不可抄襲。
- (c) 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。
- (d) 有問題者, 請發 e-mail 或 FB 私訊問助教或老師。
- (e) 不按照規定作答者, 酌量扣分。
- (f) 請參照下列文件第 2 ~ 4 頁寫作規定, 不按照規定作答者, 會扣分。

<http://www.hmwu.idv.tw/web/teaching/doc/R-how-homework.pdf>

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: dm108。密碼: xxx (上課教室號碼)。
- (c) 請上傳「學號-姓名-DM-HW3.docx」。(目錄: 「20191030-exam1」。)

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-DM-HW3-2.docx」、「學號-姓名-DM-HW3-3.docx」等等。
- (e) 上傳兩次 (含) 以上、格式不合等等酌量扣分。
- (f) 如果上傳網站出現「You can modify the html file, but please keep the link 'www.wftpservers.com' at least.」, 請將滑鼠移至「網址列」後, 按「Enter」即可。若再不行, 請換 (IE/Edge/Firefox/Chrome)。

我已經仔細閱讀上述各注意事項, 若有違背, 會自行負責。

維度縮減

1. 資料來源 (UCI): <https://archive.ics.uci.edu/ml/datasets/glass+identification>

資料檔: `glass_214x9.txt`。

Glass Identification Data Set 是玻璃識別資料集，共有 214 個觀察值，具有 9 個變數 (RI, Na, Mg, Al, Si, K, Ca, Ba, Fe) 及一個類別變數 (class, 6 個類別)。請使用 PCA {FactoMineR} 對此資料進行主成份分析。(提示: 主成份分析應包含哪些部份，請參考講義。)

2. 資料來源 (UCI): <https://archive.ics.uci.edu/ml/datasets/Wine>

資料檔: `wine_178x13.txt`。

Wine Data Set 資料集是在義大利不同地點所生產的三種類別葡萄酒資料 (此類別變數記做 y)，共 178 筆，具有 13 個變數 (特徵)(此 13 個變數記做 X)，皆為量測酒之化學成份所得到的數值。

- (a) 對此資料進行四種維度縮減方法: MDS、PCA、ISOMAP 及 SIR，並畫出維度縮減後的資料於二維平面的投影散佈圖，需用顏色標註三種酒之類別 (註: 各維度縮減方法，請自行選用合適之輸入參數)。(維度縮減講義: 105/144)
- (b) 承 (a) 小題，畫出 co-ranking 矩陣圖，並計算 LCMC($K = 7 \sim 11$)。
- (c) 以下程式碼是利用 R 套件 `e1071` 中之 SVM 分類器 (使用預設之參數)，將資料 (y, X) 建立一分類模型後，計算分類正確率之範例。(註: 此處不是預測正確率)

```
library(e1071)
attach(iris)
X <- iris[, 1:4]
y <- iris[, 5]
model <- svm(X, y)
pred <- predict(model, X)
accuracy <- sum(diag(table(pred, y)))/length(y)
accuracy
```

以 SVM 計算 Wine Data Set 之分類正確率。

- (d) 承 (b) 小題，以維度縮減後的資料 (依序使用 1 維 ~10 維) 進行 SVM 建模，並計算分類正確率。以 PCA 為例，若維度縮減後的資料變數為 $\{PCA_1, PCA_2, \dots, PCA_{10}\}$ (這裡僅取至第 10 維)，則取 $\{PCA_1\}$ 建模，計算分類正確率；再取 $\{PCA_1, PCA_2\}$ 建模，計算分類正確率，依此類推至 10 維。將結果畫成一個線圖，橫軸為「維度:1, 2, ..., 10」，縱軸為分類正確率，圖形裡 3 條線 (需為不同型式之線條) 各代表四種維度縮減方法於不同維度的 SVM 分類正確率。(維度縮減講義: 143/144)

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。