

管理數學 (107-1)

期末專案

2018-11-07; Version 2.0

授課教師: 盧信銘

本學期的期末專案將讓各位練習資料分析的工作。我們的主角是迴歸分析。迴歸分析在課本的第 4.6 節，然而基本的單變數迴歸在高中的課程中已經有介紹，所以各位同學不應該要等到 4.6 節上完之後才開始著手。期末會有許多死線，及早開始可以讓大家都較充裕的時間完成工作。

分組

自行分組，每組三至五人。如果你在期限內沒有分組，將會被隨機分組。

任務與資料集

本專案的任務為對一個選定的資料集進行迴歸分析。這個任務至少應包含以下四個部分：

- 訂定分析問題與尋找資料集。
- 描述資料集的特性。
- 迴歸分析。
- 結論與建議。

訂定分析問題與尋找資料集

在進行迴歸分析前，需決定分析的問題與所需的資料集。這兩件事可以一起決定，或先後決定。舉例而言，分析的問題可以是房價如何受交通建設，如捷運站的距離或交流道的距離所影響。資料集則可以由實價登錄與其他資料(如捷運出口位置等)製作。

分析的問題應該是有趣的、有挑戰性的。而使用的資料集必須要能回答你想要分析的問題。下面列出幾個常用的資料集來源

- UCI Machine Learning Dataset: <http://archive.ics.uci.edu/ml/index.php>
- Kaggle: <https://www.kaggle.com/>
- Taiwan Government Open Data: <https://data.gov.tw/>

如果你使用現成的資料來源，需註明資料的來源 (如提供者與網址)。如果你分析的問題與方法是參考其他人的專案，需註明來源，並說明你的結果與其分析

結果的異同。沒有適當說明者將會被視為抄襲並扣分。如果一個專案與其他專案(不論是網路上找到的或是其他同學的) **100%**相同或極為相似，這個專案將會失去大部分的分數。

除非你分析的問題非常特別，你的資料集應包含一百筆以上的資料，且應有至少四個以上(有意義)的欄位。

描述資料集的特性

分析應包含資料特性，如筆數、資料欄位、敘述統計 (包含相關係數)、統計圖表等。

迴歸分析

迴歸方法應以 **numpy** 矩陣運算自行實作，並討論迴歸係數的意義。另外應計算每筆資料的真實值與預測值的差(名為殘差)，並討論殘差絕對值較大的資料背後可能的原因。

結論與建議

列出迴歸分析的問題與結論，並探討可能的改進方向。

計劃書

各組須在指定期限內繳交計劃書。計劃書的長度為 **1 – 2 頁 A4 紙**。內容須包含：

- 組員姓名與學號。
- 組長姓名。
- 分析問題。
- 資料集來源、簡單描述 (以文字描述資料集的特性)、資料筆數、欄位數量、欄位名稱與說明、**5 至 10 筆**的範例資料。

規定

- 每組上台報告的時間為 **12 分鐘**。至少需要有兩位同學報告。所有組員都應出席。
- 期末報告的成績有 **40%**為組間互評決定，剩餘 **60%**由授課教師決定。
- 報告結束後 **24 小時**內需繳交投影片與程式 (以 **7z** 打包上傳即可)。期限內未繳交者將失去期末報告的成績中的 **40%**。
- 為了維持公平性，並讓各組選擇的主題有多樣項，本次期末報告不能使用新生專題期末專案的資料集。但方法本身，如資料視覺化，則不受此限。

評分標準

- 分析問題的新穎性與有趣的程度。
- 資料集取得的難度。
- 分析的正確性與完整性。
- 報告流暢度、精確表達、投影片美感等。

重要日期

- 2018-11-19: 分組並向助教登記最後期限
- 2018-11-26: 繳交計劃書
- 2018-12-03: 繳交第二版計劃書 (指定組別)
- 2018-12-24: 上台報告