

---

## 統計學習與深度學習期末專案 (Fall 2022)

---

2022-10-12; Version 1.0

期末專案的目標是讓每組同學以統計學習與深度學習的方法，實作自己選定的任務，並評估其效能。動手做最能增進各種技巧的學習。我們期待每組同學能 (1) 設定一個有趣且有挑戰性的任務或問題，(2) 建構、收集、整理、清理、整併資料，(3) 建構與評估模型，(4) 展示成果。

除了本文件，請參考 [Hands-On\_Machine\_Learning]中的第二章以及 Appendix B。本說明中的步驟與綱要主要來自前述兩個來源。

一個專案主要包含八個部分 (cf. Appendix B)：

1. Frame the problem and look at the big picture.
2. Get the data.
3. Explore the data to gain insights.
4. Prepare the data to better expose the underlying data patterns to Machine Learning algorithms.
5. Explore many different models and short-list the best ones.
6. Fine-tune your models and combine them into a great solution.
7. Present your solution.
8. Launch, monitor, and maintain your system.

每個專案都應該要包含前述的七大項的內容。最後一項可以忽略。與 Appendix B 以及第二章重複的部分不在此贅述。

## 分組

自行分組，每組三至六人，請取一個組名。如果你在時間 A (具體時間見重要日期小節) 前沒有分組，將會被隨機分組。

組名的選取只要不違反善良風俗，並沒有其他限制。如果你需要靈感，可以參考網路上的中文詞表(例如: <https://drive.google.com/file/d/1N-Gl1vtqZuyYtnsiK5TmVmphoN5ZStN7/view?usp=sharing>)。形容詞 + 名詞是一個方便的組合，如: 卷曲的山水、跳舞的杜鵑花。挑選一個符合主題與組員特色的組名會讓人耳目一新。

分組資訊請自行上網登記: <https://reurl.cc/MNe2Ep>

## 專案任務與建構資料集

期末專案的任務沒有限制，但應該要用到課程中所介紹的方法。除了自由發想或自行 Google 之外，我們提供幾個之前的期末報告的投影片供大家觀摩 (見 COOL 課程內容)。

專案任務與資料集是一體的兩面。這兩件事可以一起決定，或先後決定。使用的資料集必須要能幫助你處理專案任務。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集，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，並說明你的過程、結果與參考專案的異同。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。

資料集的建構很花時間，但可以由思考專案任務與資料集的關係，設計資料收集的流程與體驗實際收集資料的過程，能對於課程中所介紹的各種方法有更深入的了解。我們鼓勵各組可以針對創新的任務自行收集資料，或是合併現有的開源資料與自行建構的資料集，以達到更好的結果。任務的創新程度與資料集建構的難度為評分項目之一。

## 計劃書

各組須在時間 C 前繳交計劃書。計劃書的長度為 1–2 頁 A4 紙。內容須包含：

- 組員姓名與學號。
- 專案任務，為什麼這是有趣或有意義的任務？
- 資料集來源。是否自行建構資料集、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。

計劃書請在時限內上傳至作業區"計劃書(第一版)"位置，並至下面位置登記時間 D 各組與老師討論計畫書內容的時段：

<https://docs.google.com/spreadsheets/d/1VZvchaKVqbFOobBnliD7qWiUTIPiUmDaRx3pTi1g0Qc/edit?usp=sharing>

## 實體報告或預錄報告影片

報告的內容應包含前述機器學習專案的七大項目。每組報告的時間為 15 分鐘。由於時間有限，報告時需掌握重點。

各組同學可以自行選擇實體報告或預錄報告影片並自行上傳至 COOL。實體報告時段(F2)為上課時間。欲實體報告的組別請在時間 E3 之前向助教登記時體報告時段，助教會參考登記志願安排報告時段，於時間 E4 公布。實體報告時會錄影。錄影的內容會上傳至 COOL 作為組間互評的材料。

如果沒有登記實體報告則需以預錄報告影片的方式進行期末報告，並在時間 F1 前上傳至 COOL。錄製程式可以使用 Evercam (至 [my.ntu.edu.tw](http://my.ntu.edu.tw) 下載)或其他螢幕錄影程式如 Sharex (<https://getsharex.com/>)或 Camtasia (<https://www.techsmith.com/video-editor.html>)。Camtasia 須付費但提供一個月試用。Camtasia 有提供剪接功能。鼎鼎大名的 Open Broadcaster Software (OBS) (<https://obsproject.com/>)也可以考慮。另也可使用 Openshot (<https://www.openshot.org/>) 剪接。

未在指定時間進行實體報告且未在時限內上傳預錄報告影片者，將會被扣分(見“規定”小節)。

詳細的專案內容請整理在 Jupyter Lab 中。為了方便閱覽，請順便提供將 IPYNB Export 成 HTML 格式的檔案。投影片、IPYNB 檔案、HTML 檔案、資料檔、其他相關文件請上傳至 COOL 作業區 "期末報告檔案上傳"。大型資料檔(>500MB)應利用檔案共用服務(如 Google Drive)提供的連結分享，勿直接上傳至 COOL。

## 組間互評

所有修課同學將於時間 H 前進行組間互評。評分方式為每位同學對其他組別評分。組間互評時不需對自己組評分。評分標準後面有詳細討論。

## 組內互評

所有修課同學請於時間 H 前進行組內互評。組內互評的結果將會用來調整每個組員的成績。詳細資訊另行公布。

## 規定

- 未在時間 G 前繳交者將依遲交程度打折。遲交一個小時內期末報告成績  $\times 0.9$ ，遲交兩個小時內期末報告成績  $\times 0.8$ ，依此類推。
- 期末報告的成績由授課教師( $\times 0.7$ )、組間互評( $\times 0.3$ )、組內互評 ( $\pm 8\%$ ) 決定。舉例: "彩色氣球"組別的得分為授課教師 8.5 分、組間互評 8.9 分。"彩色氣球"組員共五位，互評結果顯示組員一的貢獻明顯低於其他組員，因此組員一調整係數為  $\times 0.96$ ，其他組員為  $\times 1.01$ 。綜合上述計算，組員一分數為  $(8.5 \times 0.7 + 8.9 \times 0.3) \times 0.96 = 8.275$ ，其他組員的分數為  $(8.5 \times 0.7 + 8.9 \times 0.3) \times 1.01 = 8.706$ 。

## 評分標準

- 計劃書 (10%)。只要準時交，有出席時間 D 之討論者即為滿分。如被指定需要繳交第二版，則在時間 E 以前繳交，並在隔日上課時間再次與老師討論計畫書內容。未完成組別會被扣分。最後做的主題與計畫書內容不同不會被扣分。
- 分析主題的新穎性與有趣的程度 (25%)。
- 資料集建構的難度 (25%)。
- 分析的正确性與完整性 (25%)。
- 報告流暢度、精確表達、投影片美感等 (25%)。

## 重要日期

- 時間 A: 2022-10-19 下午 11:59 以前: Email 分組給助教最後期限
- 時間 B: 2022-10-20 下午 11:59 以前: 助教公布分組名單
- 時間 C: 2022-10-26 下午 11:59 以前: 繳交計劃書至 COOL 作業區，並登記 10/27 計劃書實體討論時段。
- 時間 D: 2022-10-27 上午 9:10 至中午 12:20: 各組與老師討論計畫書內容。
- 時間 E: 2022-11-02 下午 11:59 以前: 繳交第二版計劃書 (指定組別) 至 COOL 作業區。
- 時間 E2: 2022-11-03 上午 9:10 至中午 12:20: 各組與老師討論計畫書內容 (指定組別)。
- 時間 E3: 2022-12-12 下午 11:59 以前: 登記實體報告時段志願。
- 時間 E4: 2022-12-13 下午 11:59 以前: 公布實體報告時段。
- 時間 F1: 2022-12-15 下午 11:59 以前: 自行錄製報告影片上傳最後期限 (包含報告影片、報告投影片、IPYNB 檔案與資料檔至 COOL 作業區)
- 時間 F2: 2022-12-15 上午 9:10 至中午 12:20: 課堂實體報告時段
- 時間 G: 2022-12-15 下午 11:59 以前: 實體報告組別上傳專案檔案期限 (包含報告投影片、IPYNB 檔案與資料檔至 COOL 作業區)
- 時間 H: 2022-12-19 上午 9:00 至 12/22 下午 11:59: 觀看各組影片，並進行組間互評與組內互評。