

建模與預測 HW01 作業說明

目標

必須要繳交的

報告(80 %)

模型(20%)

挑選特徵紀錄

程式碼

最後上傳

其他

額外加分題(+10%)

補充資料

目標



血液透析：14 天後 HB 範圍是否在 11-12 之間且變化幅度在正負0.5之間(最佳化)？

→ 二分類問題

- 將模型訓練至準確度越高越好+挑出最好的模型 → 可能的方向
 - 使用特徵挑選方法
 - 超參數最佳化
 - 選擇不同的模型
- 嘗試不同的模型以及比較
 - 選擇適合的模型、了解其參數調整用意
 - 能算出評估指標與圖表
 - Model Evaluation
 - Cross Validation Result
 - 圖

必須要繳交的

報告(80 %)

- 一律接受 pdf 檔
- 需要的內容：
 - 建模訓練過程記錄
 - 如特徵挑選、參數最佳化、資料切分方式、任何你做的步驟都可以紀錄。
 - 兩種指定模型
 - SVC (scikit learn)
 - 參考 1
 - KNeighborsClassifier (scikit learn)
 - 參考 1
 - 參考 2
 - 其他模型”們” (加分題，可以很多種。+5%/每個)
 - 需要有的指標圖表 (可參考 10/26 課堂示範之程式碼)
 - Model Evaluation
 - Accuracy, Specificity, Sensitivity, MCC (Matthew Correlation Coefficient), PPV, NPV
 - Cross Validation Result
 - Accuracy, Precision, Sensitivity, Specificity, MCC
 - 圖
 - Confusion matrix
 - ROC curve (各模型比較，最少也會有 SVC + KNeighbors)
 - 上傳的模型名稱以及其對應的訓練紀錄

- 程式碼及對應的實驗部分
- 結論
 - 為何挑選這個模型？
 - 根據哪些指標或圖？
 - 實作中發現了哪些問題？(加分題，+5%)

模型(20%)

選擇一個最終模型給我們助教作 testing，請在report中紀錄此模型是哪一次訓練過程

挑選特徵紀錄

如果你的模型有挑選特徵，請寫一個 csv 格式純文字檔，如以下範例：

```
d3, d20, d22, d37, c46, c69, c85
```

程式碼

格式不拘(py, ipynb)、檔案數量不拘，但必須要在報告中說明清楚在哪裡使用了這個程式

最後上傳



請最後把 報告、模型、(挑選特徵紀錄)、程式碼 壓縮成 .zip 檔並命名 [學號]_HW01.zip (ex: 310350035_HW01.zip)

其他

額外加分題(+10%)



如何調整 threshold 得到不同的 confusion matrix 數值 (tp,fp,tn,fn) 以換算 sensitivity, specificity 等數值？

請寫一隻程式 (.py, .ipynb 皆可)

補充資料

[pandas 教學](#)

[scikit-learn 官網](#)