

機器學習導論  
HW2

1. 讀取 hw\_data1.csv 這筆資料，並依下列條件整理數據。(35%)
  - (a)將資料中的 Team、Yellow Cards、Red Cards 獨立取出並合併。
  - (b)承上，對此表格依照 Red Cards、Yellow Cards 數量進行排序(先紅後黃)，輸出仍為一個表格。
  - (c)計算每個球隊的到黃牌數量之平均值。
  - (d)找出進球數(Goals)超過 5 球的球隊及其數據並依進球數排序。
  - (e)取出以字母 S 開頭的球隊及其數據。
  - (f)選取除了最後 5 個 feature 外之其他 feature
  - (g)找出 France、Poland、Spain 之射正率(Shooting Accuracy)並進行排序
2. 使用 sklearn 之 Boston dataset(sklearn.datasets.load\_boston)並隨機選取此筆資料的 30% 做為測試資料，並手刻 Naïve Bayes Classifier 進行，(使用 sklearn 之模型套件此題會沒有分數)  
預測(高房價低房價)，並算出預測之準確率(accuracy)。
  - (a)適當的整理資料並將 target 之連續值依平均數分為高房價及低房價兩種類別。(10%)
  - (b)使用其他假設之 pdf 並任取兩種 feature 進行預測。(10%)
  - (c)承上，取用四種 feature 進行預測。(10%)
  - (d)取用與(c)相同之 feature，設計適當方法使 Class1、Class2 的每個 feature 共用同一個 variance 並進行預測。(20%)
  - (e)可視化各種做法之結果，並分析各種方法之結果差異，敘述你的看法。(15%)

註：繳交項目：

1. 請將所有程式碼整理成一個檔案上傳
  2. 一份 word 的作業報告，解釋結果，形式自由發揮。
- 有任何問題歡迎到 ESS305 詢問