

# 機器學習作業指示 - Homework #2

## 目標：

透過自己實現之Linear Discriminant Analysis (LDA)演算法對鳶尾花資料(Iris data set)進行分類；並利用ROC曲線和AUC觀察調整懲罰權重對模型表現所造成的影響；最後，進一步測試多類別分類策略 ( e.g. One against one strategy ) 。

## 作業內容：

### (1) 分類器實現

1. 利用程式實現LDA分類器 ( 請勿直接使用開源的LDA套件 )

### (2) 利用2-fold cross validation推估LDA在二元分類之分類率

將Iris data set的變色鳶尾 ( Versicolor ) 以及維吉尼亞鳶尾 ( Virginica ) 分別設為positive class與negative class，每個class皆採用第三種與第四種特徵。

- Step1. 將positive與negative類別的前25筆data設為training data，剩餘的50筆設為test data。
- Step2. 將training data送至LDA中訓練，並將懲罰權重皆設為1，並求出hyperplane之 weight vector及bias，並記錄下來 ( 四捨五入至小數點後第2位 )。
- Step3. 將test data輸入至Step2所訓練的LDA decision function，對50筆test data分別進行分類決策，並記錄分類率 ( 四捨五入至小數點後第2位，in % )。
- Step4. 將Step2中的training data改成test data，test data改成training data，重複Step2 ~ Step4進行交叉驗證。
- Step5. 將Step2和Step4求出的兩個分類率平均，得到平均分類率並記錄 ( 四捨五入至小數點後第2位，in % )。

### (3) 繪製ROC與計算AUC

以第三類別當作positive class，第二類別當作negative class。調整LDA的C1/C2比值，

- Step1. 繪製出在四個特徵都使用下的ROC曲線 ( receiver operating characteristic curve )，並計算其AUC ( area under the curve )。
- Step2. 僅使用第一和第二個特徵，繪製出在使用此兩個特徵下的ROC曲線，並計算其AUC。
- Step3. 僅使用第三和第四個特徵，繪製出在使用此兩個特徵下的ROC曲線，並計算其AUC。
- Step4. 討論上述三個不同情形的AUC，並解釋特徵組合和AUC之關聯性為何，AUC是否可以當作特徵比較之量化工具？

### (4) 多類別分類問題 - One against one strategy

- Step1. 針對Iris data set的山鳶尾 ( Setosa )、變色鳶尾 ( Versicolor ) 以及維吉尼亞鳶尾

( Virginica ) 分別設為Class1,2,3三個類別；每個class皆採用第三種與第四種特徵，每個class前25筆data設為training data，剩餘的75筆設為test data。

- Step2. 採用one against one strategy將LDA推廣為多類別分類器，並將懲罰權重皆設為1，根據Part2的流程( Step1 ~ Step2 )，分別建立兩兩一組的分類模型：Class1 vs. Class2，Class1 vs. Class3，Class2 vs. Class3。
- Step3. 將test data輸入至Step2所訓練的3組LDA decision function，對75筆test data分別進行分類決策，並以多數決來做為分類依據( 若同票則該測試資料判定為分類錯誤 )，並記錄分類率 ( 四捨五入至小數點後第2位，in % )。
- Step4. 將Step1中的training data改成test data，test data改成training data，重複Step2 ~ Step3進行交叉驗證，並記錄分類率 ( 四捨五入至小數點後第2位，in % )。
- Step5. 將Step3與Step4求出的兩個分類率平均，得到平均分類率並記錄 ( 四捨五入至小數點後第2位，in % )。

### (5) 編輯書面報告

1. 將上列所記錄的結果 ( 包含但不僅限於畫底線部分 ) 詳細且清楚地以表格方式呈現在書面報告中。
2. 並且針對分析方式、結果等做詳細的討論。

**作業繳交注意事項 ( 遲交一周該次作業打8折，遲繳超過一周視同該次作業0分)：**

1. 作業報告請以書面(Pdf or Markdown)呈現，並將程式碼一併壓縮在一個壓縮檔中。
2. 壓縮檔名請符合下列格式：「作業\_學號\_姓名」(如：HW2\_09511001\_王小明)，否則作業成績打8折。
3. 請將壓縮檔上傳至E3數位教學平台。

**Deadline：2024/10/15 11:59 P.M. (Two weeks from now)**