

1131資訊視覺化\_第九組

# 臺灣各縣市交通事故記錄之視覺化呈現

## — 論文報告

111304019 統計三 林承佑

107703027 資科四 崔賢燮

***Interactive Dimensionality Reduction  
for Comparative Analysis***

## 研究介紹

### *Interactive Dimensionality Reduction for Comparative Analysis*

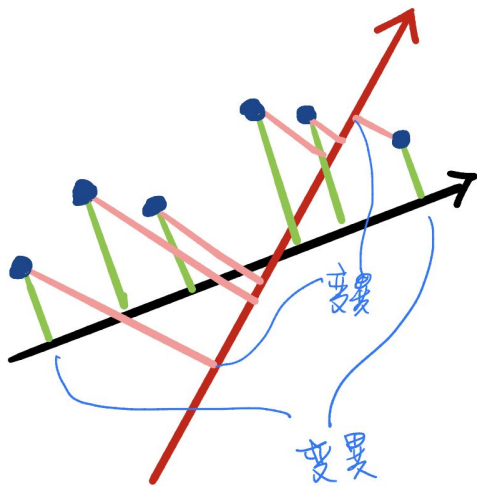
- **研究目標** — ULCA : unified linear comparative analysis
  - 一種降維方法, 找到一組線性組合, 組合間相關性為零Linear indepdently
  - 目標是在保持數據主要結構的同時消除冗餘信息。
  - 結合PCA、cPCA、LDA的主要概念並優化
    - 以致可以同時處理連續和類別的資料內容

## 研究內容和技術

- PCA

- 將高維數據嵌入到較低維的空間中，同時保留數據中最大程度的變異性
- 透過EVD得到投影矩陣  $M_{PCA}$ ，而 ULCA 在此基礎上通過特徵值縮放獲得完全去相關的結果：

$$\max_{M^T M = I_{d'}} \text{tr}(M^T C M)$$



## 研究內容和技術

- cPCA

- PCA 的一種變體，用於對比兩組數據（目標組和背景組），以發現對目標組特有的顯著特徵：
- 找到區分類別的方向，而 ULCA 在這些方向上進一步消除類別之間的相關性。

$$\max_{M^{\top}M=I_{d'}} \text{tr}(M^{\top}(C_{\text{tg}} - \alpha C_{\text{bg}})M)$$

C: 目標組和背景組的covariance matrix

X: 目標組和背景組的數據

alpha: 對比參數，**控制目標高變異和  
背景低變異的 權衡**

## 研究內容和技術

- LDA

- 類別信息, 找出能使不同群組之間的區分最大化的嵌入空間

$$\max_{M^T M = I_{d'}} \frac{\text{tr}(M^T C_{\text{bw}} M)}{\text{tr}(M^T C_{\text{wi}} M)}$$

C: 組間和組內的covariance matrix

$X_i$ : 目標組和背景組的數據點(因為是類別)

最小化組內變異性  $C_{\text{wi}}$  並最大化組間分離

$C_{\text{bw}}$

## 研究內容和技術

- **ULCA**
  - 是一個框架，統一並擴展了包括 PCA、cPCA 和 LDA 在內的多種線性降維方法
    - 廣義 cPCA (gcPCA)
    - Trace-Ratio Form

## 研究內容和技術

- 廣義 cPCA (gcPCA)
  - gcPCA 將 cPCA 的對比目標擴展為多個群組  $C_j$ 。
  - 把每個 covariance matrix 展開
    - 加入:  $\delta_{jy_i}$ 
      - $y_i = j$  時  $\delta = 1$  ; 否則為 0
      - 標記是否是屬於該群組  $j$

$$C_{wi}^j = \left( \sum_{i=1}^n \delta_{jy_i} \right)^{-1} \sum_{i=1}^n \delta_{jy_i} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^\top$$



## 研究內容和技術

$$\max_{M^{\top}M=I_{d'}} \text{tr}(M^{\top}(C_{\text{tg}} - \alpha C_{\text{bg}})M)$$

- 廣義 cPCA (gcPCA)
  - 最大化目標組(權重  $w_{\text{tg}}$ )並最小化背景組(權重  $w_{\text{bg}}$ ) 的差值
    - 每個權重就代表他對那個j群組的貢獻(0~1)
  - $C = 2$  時, 就變成原本的 cPCA的結果(  $w_{\text{tg}}=(1,0)$  or  $w_{\text{bg}}(0,1)$  )

$$\max_{M^{\top}M=I_{d'}} \text{tr} \left( M^{\top} \left( \sum_{j=1}^c w_{\text{tg}}^j C_{\text{wi}}^j - \alpha \sum_{j=1}^c w_{\text{bg}}^j C_{\text{wi}}^j \right) M \right)$$

## 研究內容和技術

- Trace-Ratio Form
  - 統一 cPCA/gcPCA 和 LDA -> ratio
  - 最大化目標組與背景組的covariance matrix
- 可以找到 equivalent的 alpha
  - 對比參數, 控制目標高變異和
  - 背景低變異的 權衡

$$\max_{M^T M = I_{d'}} \frac{\text{tr}(M^T C_{\text{tg}} M)}{\text{tr}(M^T C_{\text{bg}} M)}$$

$$\max_{M^T M = I_{d'}} \frac{\text{tr} \left( M^T \left( \sum_{j=1}^c w_{\text{tg}}^j C_{\text{wi}}^j \right) M \right)}{\text{tr} \left( M^T \left( \sum_{j=1}^c w_{\text{bg}}^j C_{\text{wi}}^j \right) M \right)}$$

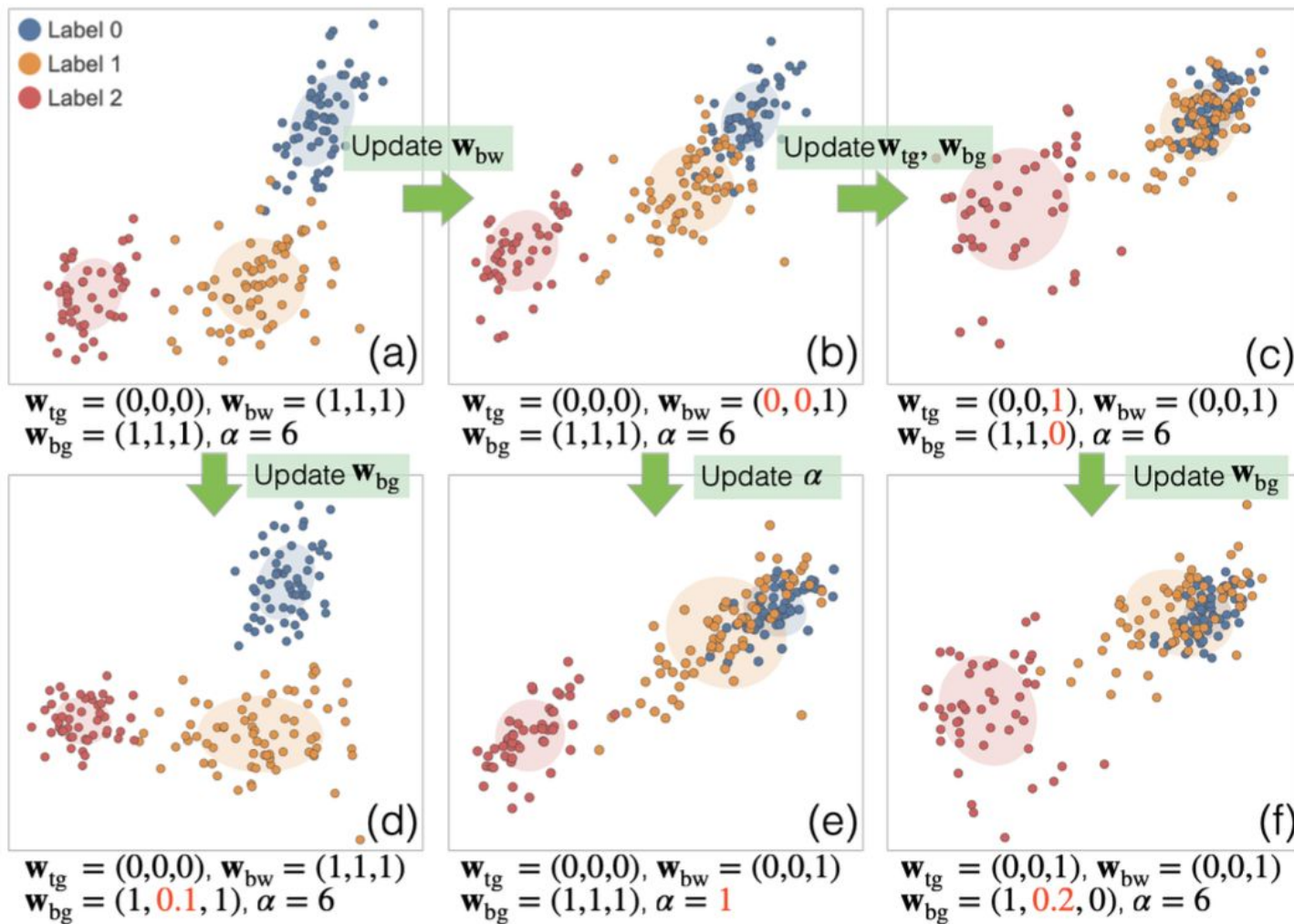
## 研究內容和技術

- ULCA
  - 從數學角度的計算和推倒，其實可以得到：
  - 用 gcPCA 擴展 cPCA 的應用範圍，處理多個群組，更精細的控制。
  - Trace-Ratio 表達為統一的優化框架
    - 最後再加入gamma Id -> Regularization 避免C為零
    - (假設gamma0=1 或 gamma1 = 1)

$$C_0 = \sum_{j=1}^c w_{\text{tg}}^j C_{\text{wi}}^j + \sum_{j=1}^c w_{\text{bw}}^j C_{\text{bw}}^j + \gamma_0 I_d,$$

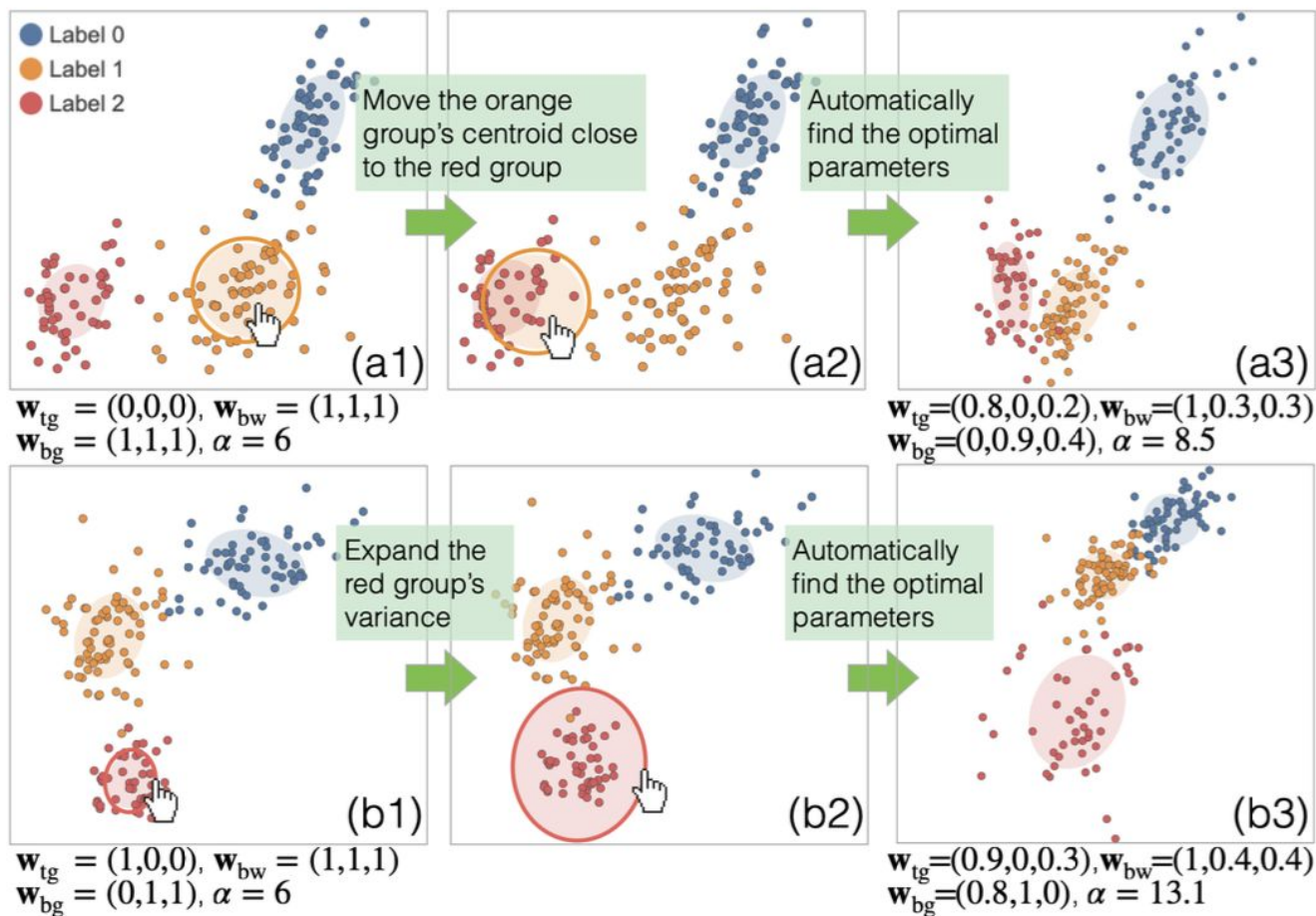
## 視覺化的實現

- $w_{tg}$  :調整目標群組的 內部變異
- $w_{bg}$  :控制背景群組的 內部變異
- $w_{bw}$  :調整群組之間的距離
- $\alpha$ :控制嵌入結果中背景變異的減少量。



## 反向參數

- **Wbg**
  - 移動群組centroid:
    - 使用者調整群組之間的**距離關係**。
  - 縮放群組的變異 (variance scaling):
    - 反映使用者改變**群組之間的相對變異程度**。(背景大小)
  - **向後權重 Backward**, 主要處理類別數據通過從數據矩陣中**移除組內平均值**, 保證去相關的變換在類別內仍保持正確的結構。



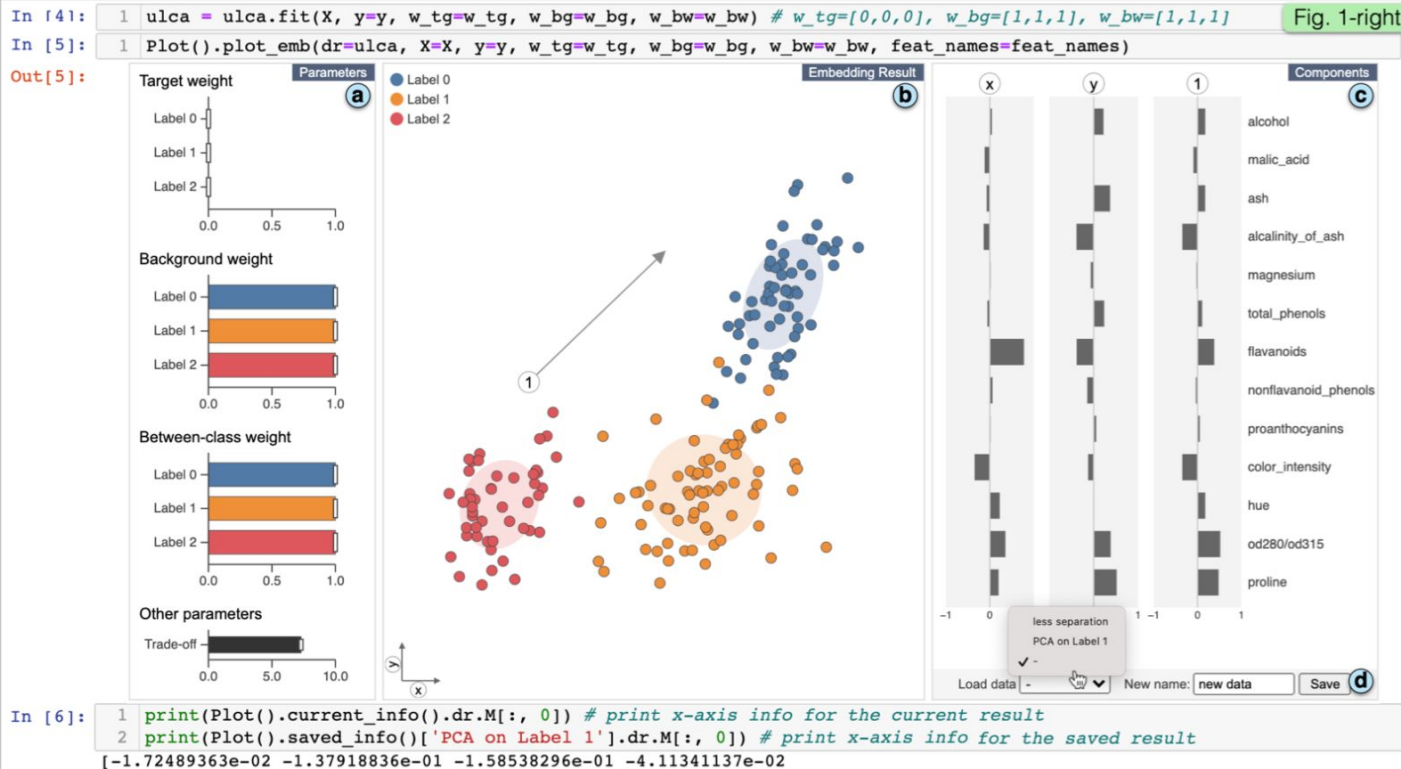
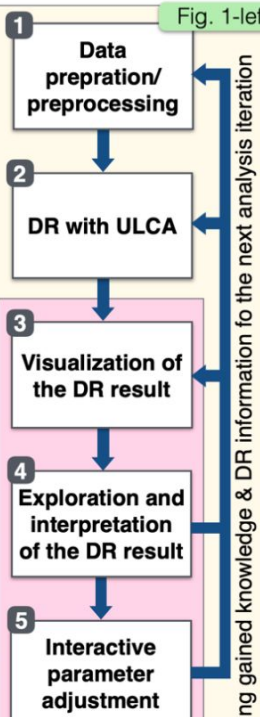
D

Fig. 5: Backward parameter selection on the Wine dataset when moving a group centroid (a1–a3) and scaling a confidence ellipse (b1–b3).

# 操作呈現

Analysis steps performed with Python scripts in the Jupyter Notebook

Steps performed with the visual interface

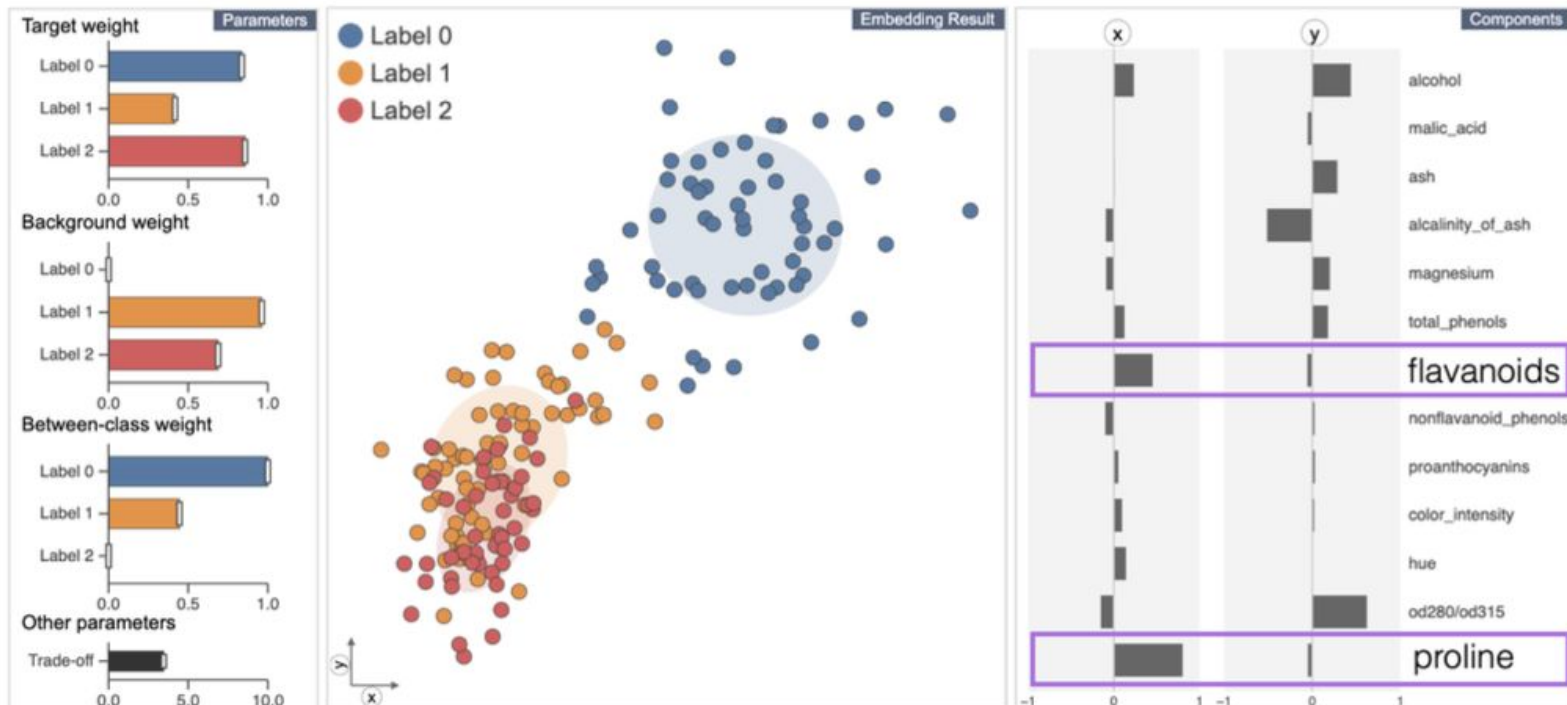




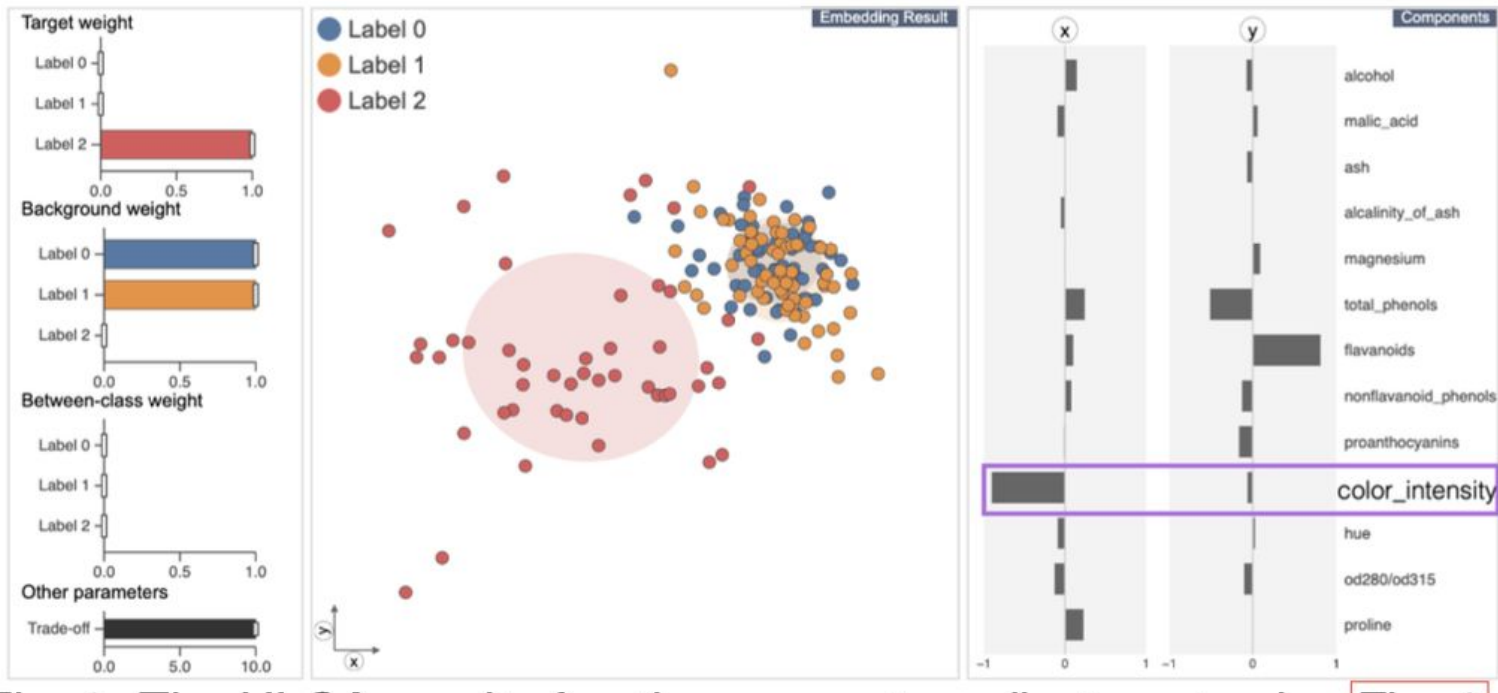
## 操作呈現

- 發現：
  - 使用者其實不需要調整alpha值，由數據的分布由程式進行計算
  - 只要調整數據的分布 = 調整目標權重
    - -> 調整目標權重就會改變背景權重
    - -> 也就影響群組之間的距離關係 即反向權重
  - 從右圖可以如同cPCA的特性與計算，得到較高的分數即為重要參數

# 操作呈現



# 操作呈現

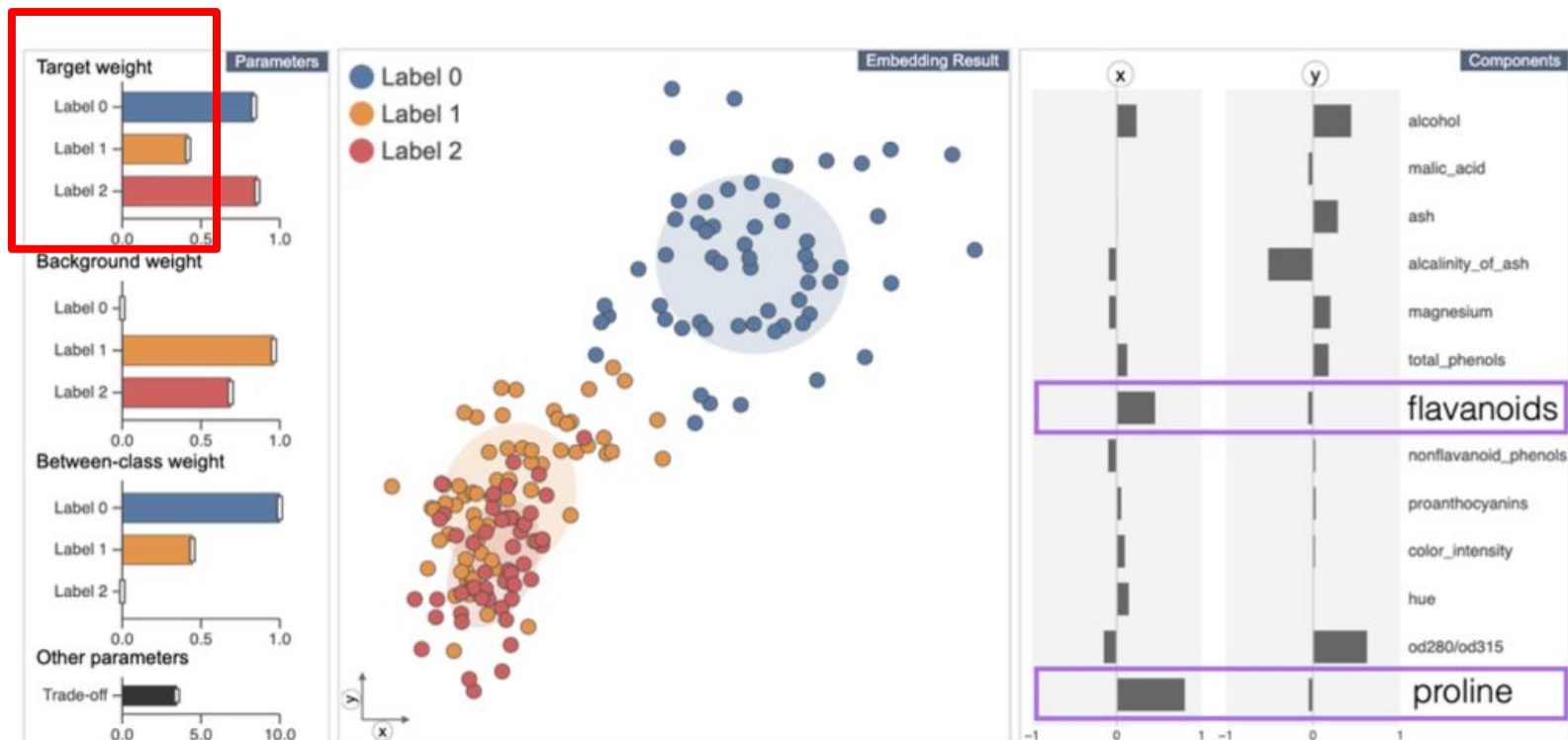


## 跟期末專案關聯

- 目標：
  - 發現交通事故數據中的潛在模式和趨勢。
  - 找出一般事故發生的時間段、地區、事故類型以及肇因因素。
- 應用：
  - **淺在模式** -> 由降維度的方式，保留最主要的變數
  - 由使用者進行篩選，而非我們先做篩選
    - 呈現的模式與內容選擇權還給使用者

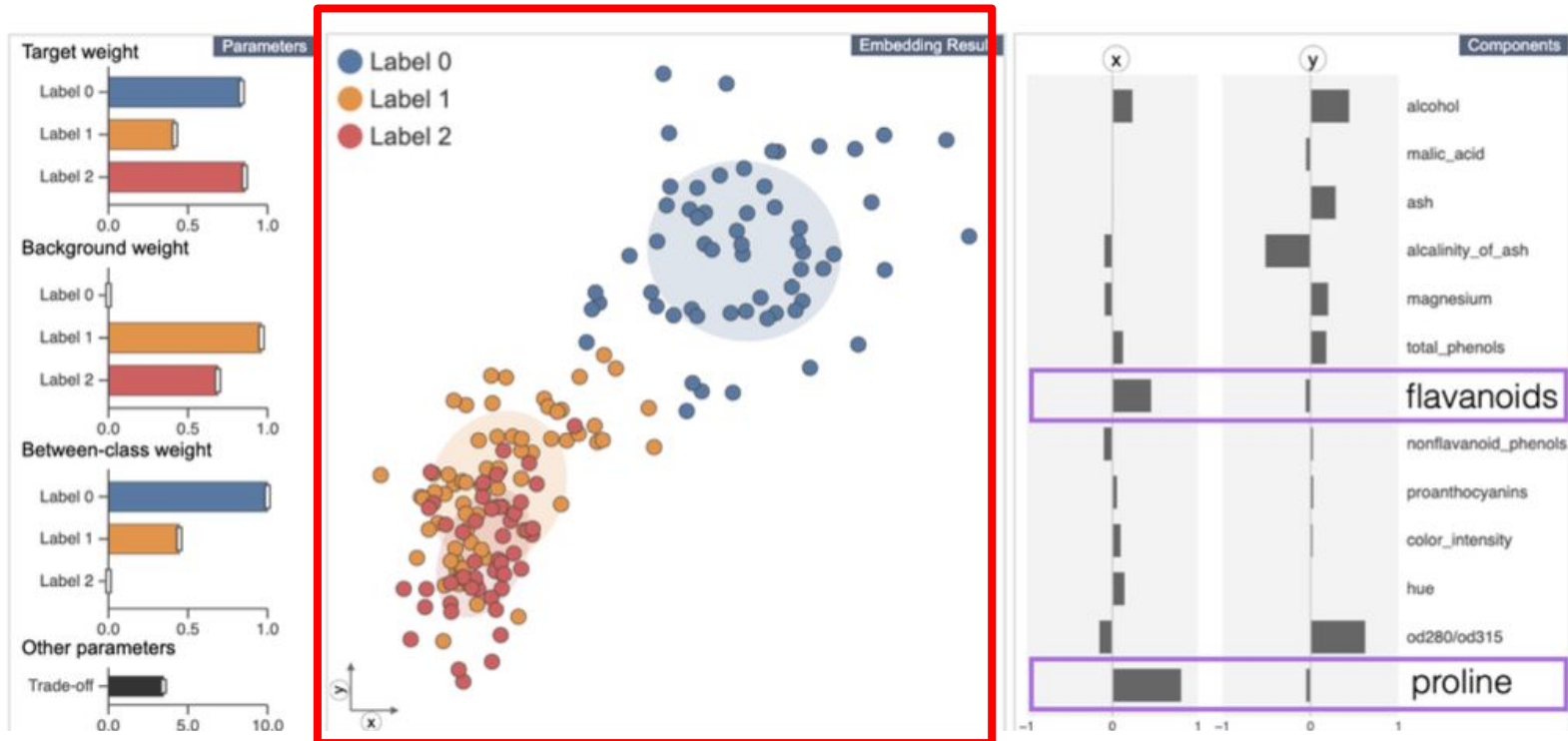
## 跟期末專案關聯

目標項目，可以由使用者選擇他在意的：  
例如：自撞、撞樹、撞人



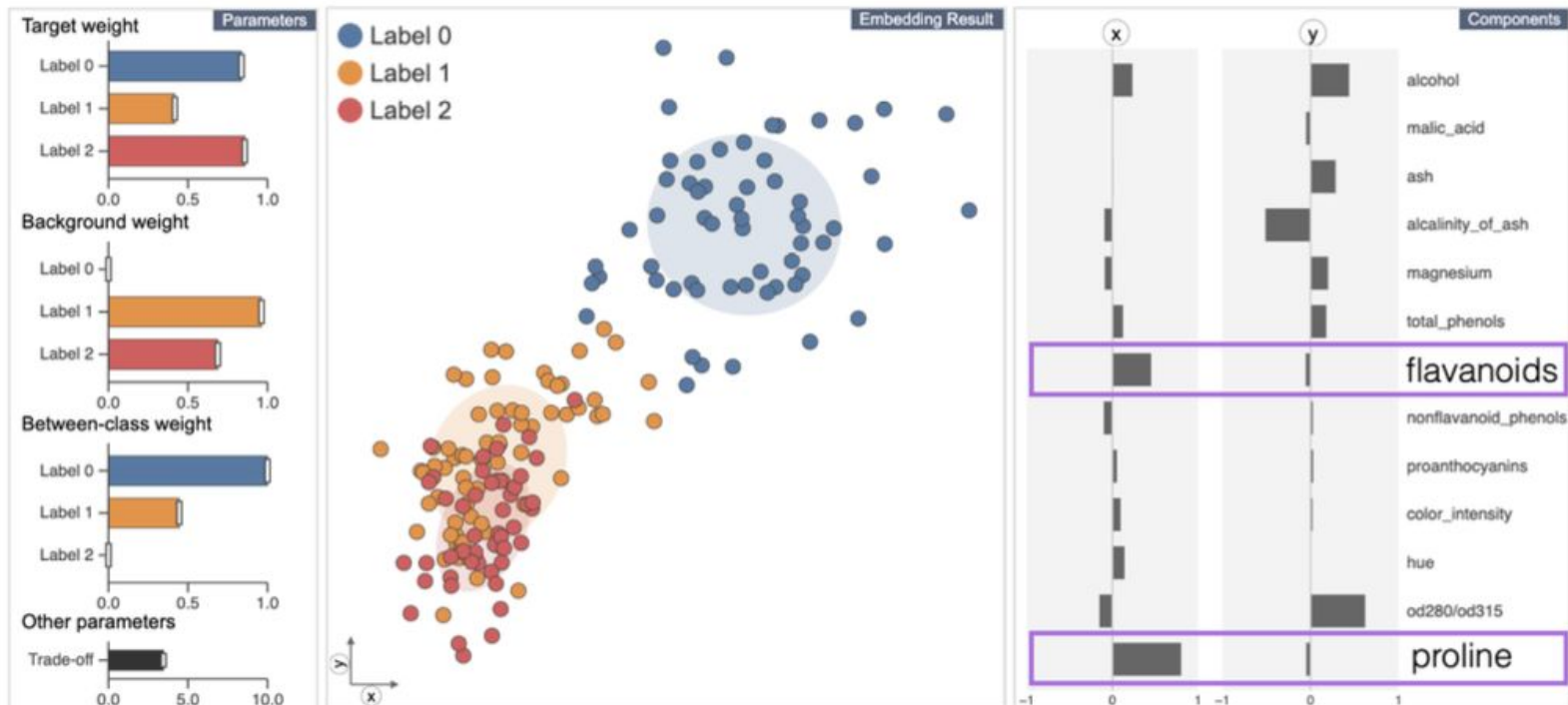
## 跟期末專案關聯

只要調整中間的數據分布或範圍大小，  
就可以從右側得到重要參數！



## 跟期末專案關聯

不再由我們先處理數據可以呈現的內容  
而藉由固定少數的數量去呈現



*The END*

[作者的Demo](#)