

# 人工智慧期末專題報告

## [題目]紅酒品質

### 1. 組員：

11160924-林鈺軒

11160801-陳禹豪

11160605-蔡承哲

11160534-許友銓

### 2. 主題：

分類

### 3. 資料集：

Wine Quality Data Set

名稱：

Winequality-red.csv

來源：

<https://archive.ics.uci.edu/ml/index.php>

### 4. 資料集介紹分析

本資料集中的數據包含 1599 個紅酒樣本和 13 個分類（固定酸度、揮發性酸度、檸檬酸、殘糖、氯化物、遊離二氧化硫、二氧化硫總量、密度、pH、硫酸鹽、酒精、評價）。

### 5. 目標

我們想透過紅酒的數據，以科學的方式分析紅酒品質

### 6. 解釋採用的模型(Model):至少 2 個模型+調整參數

**SMoreg**：支持向量機迴歸它使用核技法(kernel trick)將非線性呈現的資料轉換成容易區隔的分佈。支持向量機演算法的特色是計算速度快、預測成效也挺不錯的。

**Lazy. Ibk.**：最臨近分類器，根據案例的最靠近距離個數來決定分類

## 7. 解釋結果/結果比較

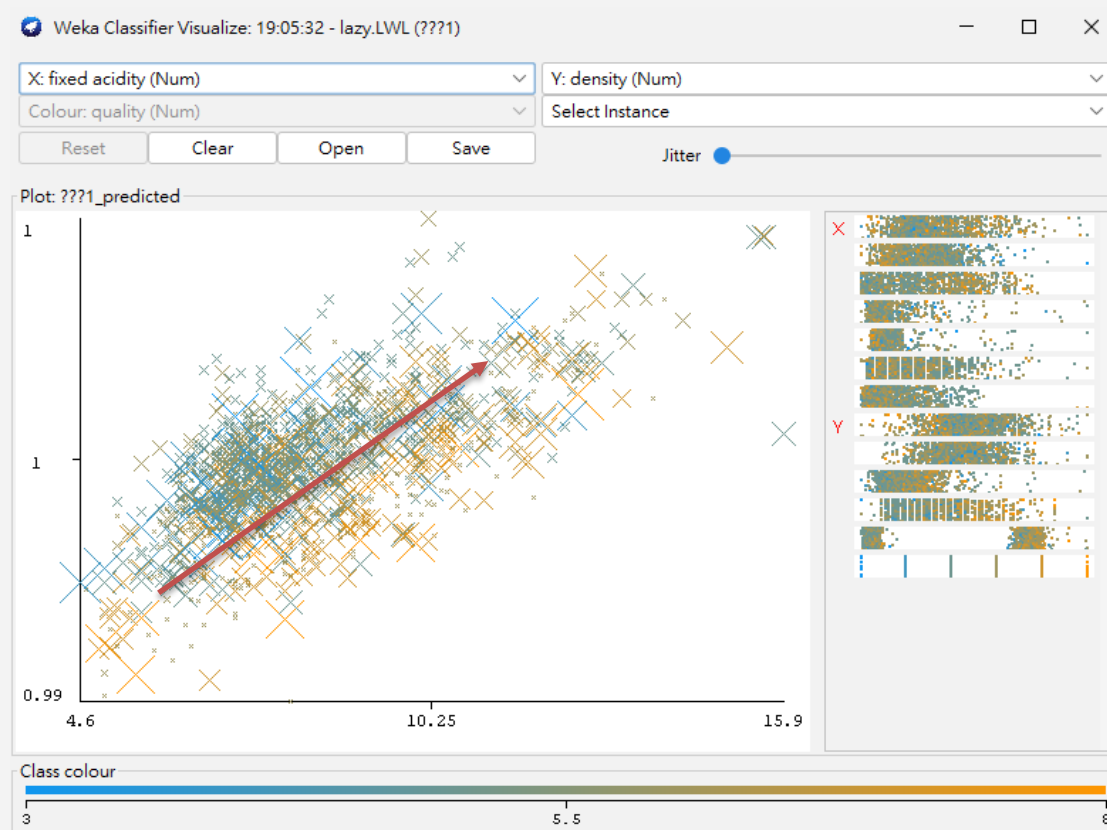
我們認為，品酒師在品定紅酒等級時，應該是不會被資料中大部分的分類所影響的，通過紅酒內容物的數據以及紅酒品質的關聯性進行比較就能發現，在各分類下比較都很難有相關程度。

但依據我們的分析，以數據當中能刺激感官的分類進行對應分析，我們發現檸檬酸(citric acid)、揮發性酸度(volatile acidity)、酒精(alcohol)這三個項目比較能夠刺激品酒師感官，讓品酒師依據這些刺激定下紅酒的等級。

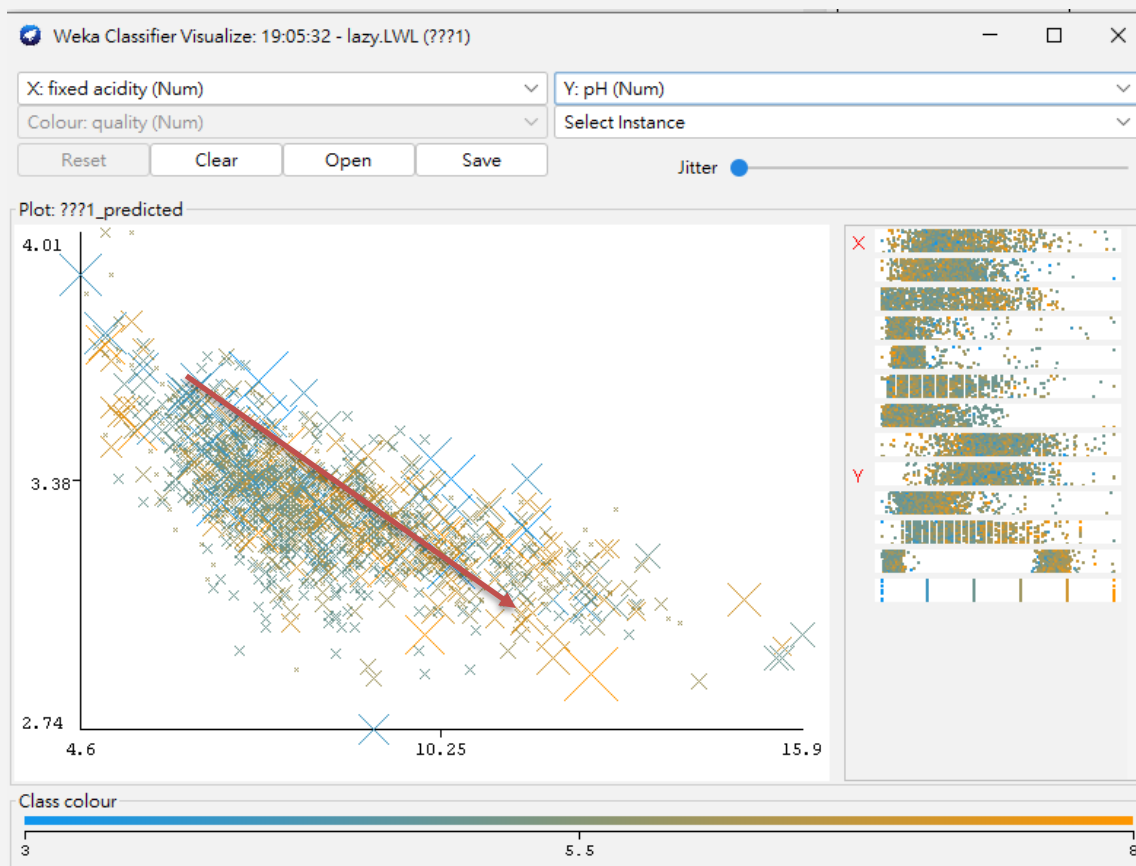
我們發現，固定酸度和檸檬酸高度負相關，但是與揮發性酸度的關聯性卻相對較弱，而紅酒中的檸檬酸、固定酸度、揮發性酸度卻關聯度較高，檸檬酸與固定酸度、揮發性酸度負相關關係，而且關聯度也都相差不多。

我們也發現，固定酸度和PH值是高度負相關的(如下圖)，其相關度是最高的，固定酸度的增加會使得PH值降低，偏向酸度。

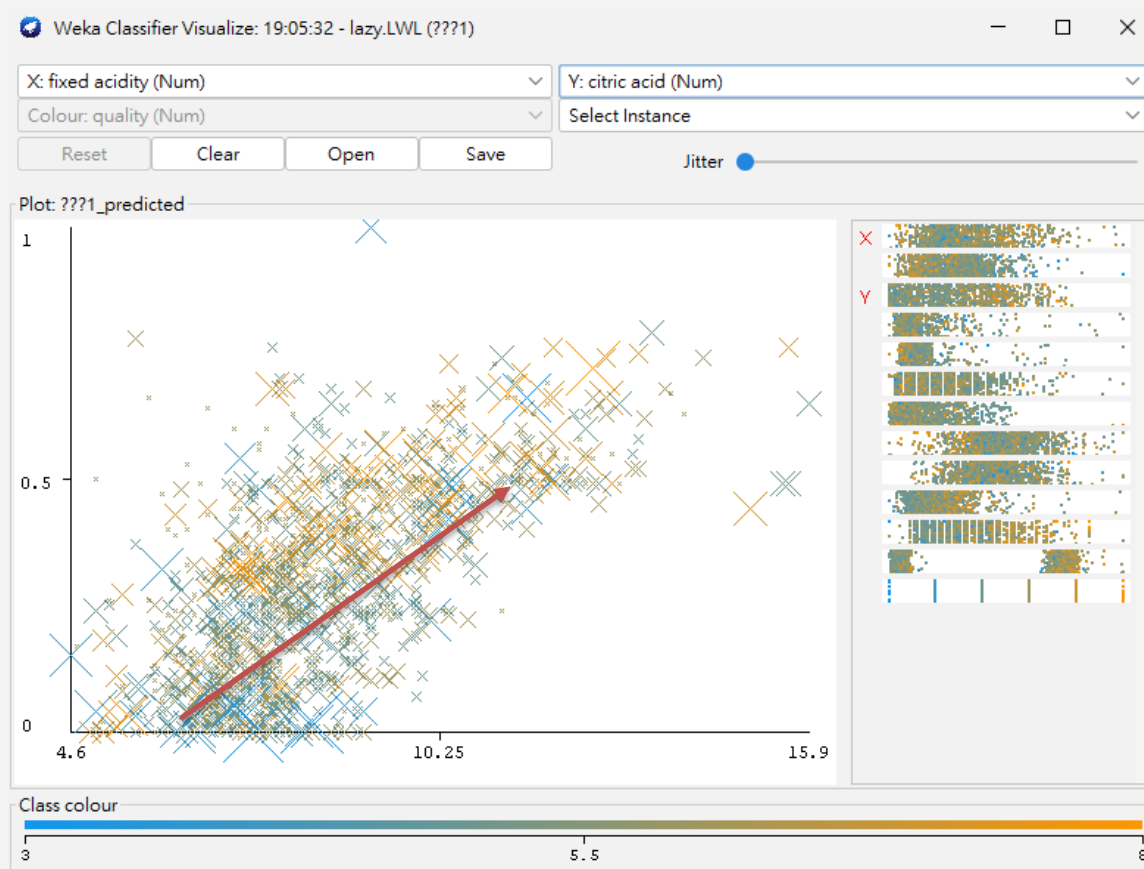
數據展示：



固定酸度和密度(正相關)



固定酸度和 PH 值(高度負相關)



檸檬酸和固定酸度呈高度的(正相關)

## 8. 總結

通過以上分析，我們可以總結說，**檸檬酸、酒精含量、揮發性酸度**這三個分類可以直接影響對紅酒的感覺，也就是說，這三個類別可以直接影響品酒師的感受，來評斷一隻酒的等級。

通過這個結論，我們認為，如果在紅酒釀造過程中，**對這三個項目加以控制**，例如添加檸檬酸或者加入檸檬汁，或者增加酒精含量等，我們就有方法可以提高獲得高評價的酒的機率。

## 9. 心得

陳禹豪:透過這次的報告，使我們更熟悉 weka 的功能，並用它進行各種數據分析，在紅葡萄酒評級的應用中，我們可以利用 Weka 來分析大量的紅葡萄酒資料，並建立模型來預測紅葡萄酒的品質。例如，我們可以使用 Weka 來分析探討酒中不同的成分，例如酒精濃度還有酸度等因素對紅葡萄酒品質的影響。我們還可以利用 Weka 來探索葡萄酒中不同成分之間的相互作用，並研究它們對紅葡萄酒品質的貢獻。通過使用 Weka 進行數據分析，我們可以得出有助於提升紅葡萄酒品質的結論，並讓我們在未來的紅葡萄酒製作過程中，有很大的幫助。

林鈺軒:經過這次的專題實驗課，我在當中學到了，Weka 這個應用軟體，是可以將大量數據透過機器學習，來產生我們所想要分析的數值的，在小組當中，我們利用了「分類」這項功能，來分析 UCI 當中所提供給機器學習的數據，也從中或獲得了許多，相信未來如果能利用程式碼來自己編寫出機器學習的程式，會相較這次的經驗更有收穫。

蔡承哲:當初上課時學到分類的時候，還不是很了解分類的意義，實習課做的作業也是在同學的教導做的，沒有自己去思考分類是甚麼，而這次分組透過實作和組員的討論，更加了解分類這項功能，也對 Weka 的程式更加了解，相信在未來的專題中，也會使用這次所學的來完成專題。

許友銓:在這次的專題中我們用到了 weka 的分類功能，除了讓我熟練 weka 上的功能和操作方式外，也更清楚分類的涵義以及應用，像是可

以用來分析大量數據，並且可以得到明確的結果，在之後的求學過程中可以應用本次所學到的東西。