

教案內容架構圖

最佳化理論之應用

課程
單元

迴歸分析與分類系統
(基礎)

迴歸分析與分類系統(進階)

萃取隱藏的特徵

目標

目前許多大數據分析或機器學習應用(例如:垃圾郵件篩建、信用卡核發、用電預測等)皆可利用迴歸分析與分類系統完成,本單元主要教授這些『基礎知識』及其相關應用,目標將令學生先了解這些應用背後的基礎。

基於前5週的基礎技巧,將更深入討論這些技巧的優缺點,針對這些缺點,將提供其進階的技巧,包含如何『嵌入大量的特徵』及『融合預測性的特徵』。

之前所教授之技巧接假授對機械模型據初步了解,接下來的幾門課我們將教授如何從數據中萃取更大量的特徵

時間

4-5 週

4-5 週

4-5 週

教學
資源

1. 上課講義
2. 程式撰寫
3. 實例操作

4. 上課講義
5. 程式撰寫
6. 實例操作

7. 上課講義
8. 程式撰寫
9. 實例操作

主要
教學
活動

1. 線性迴歸(老師授課)
2. 非線性迴歸(老師授課)
3. 分類系統(老師授課)
4. 參數及模型選擇(老師授課)
5. 漫想~應用機器學習(上述方法)於生活上的應用

1. 線性支持向量機(老師授課)
2. 對偶支持向量機(老師授課)
3. 核型支持向量機(老師授課)
4. 軟式支持向量機(老師授課)
5. 核型羅吉斯迴歸(老師授課)
6. 支持向量迴歸(老師授課)
7. 應用上述方法於生活上的應用

1. 參數評估(老師授課)
2. 數值最佳化(老師授課)
3. 叢集(老師授課)
4. 矩陣分解(老師授課)
5. 應用上述方法於生活上的應用

評量

實例操作、習題演練、專題製作

單元課程設計

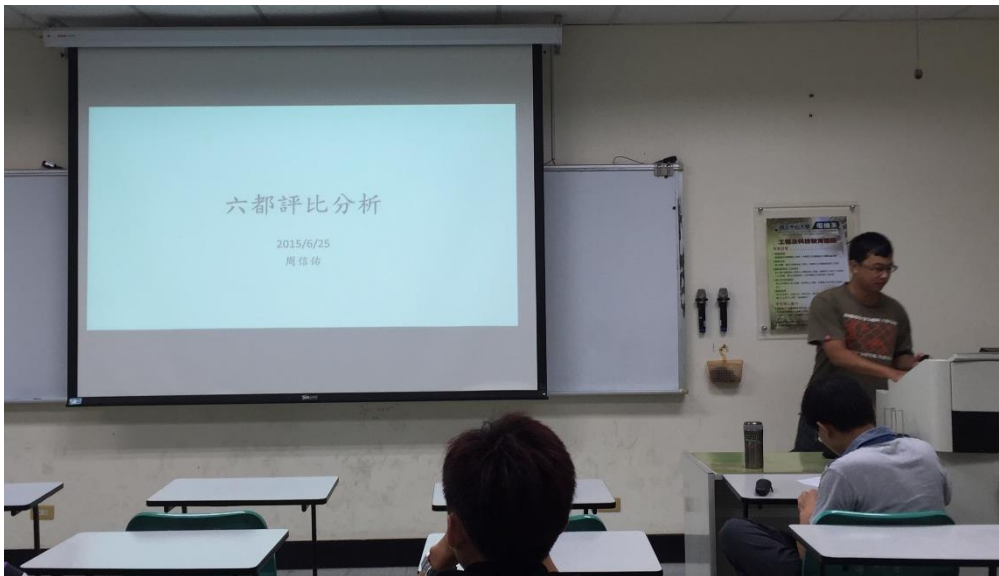
教學主題	迴歸分析與分類系統	設計者	中山大學 溫朝凱			
教學對象	碩士生及大學部高年級學生	教學時數	4-5 週			
教材來源	1. Learning from Data: A Short Course, Abu-Mostafa, Magdon-Ismael, Lin, 2013. 2. An Introduction to Statistical Learning, James, Witten, Hastie Tibshirani, 2010					
教學重點	活動 3: 萃取隱藏的特徵 1. 參數評估(老師授課) 2. 數值最佳化(老師授課) 3. 叢集(老師授課) 4. 矩陣分解(老師授課) 5. 漫想~應用機器學習(上述方法)於生活上可能的應用					
學習目標	目前許多大數據分析或機器學習應用(例如:垃圾郵件篩建、信用卡核發、電能預測等)皆可利用迴歸分析與分類系統完成,本單元主要教授這些『基礎知識』及其相關應用,目標將令學生先了解這些應用背後的基礎(具體而微)。					
科學想像力教學目標	核心能力	具體目標				
	漫想	●能從課本生活例子中發現生活周遭的不便或待解決的問題				
	聯想	●能思考、討論,想出更多解決問題的點子。				
	奇想	●在眾多點子間引起共鳴,利用重組形成新奇的點子。				
	妙想	●能將創意點子精製化,繪出設計圖,內容涵蓋功能、材料等,並付諸實行。				
節次	教學流程及內容設計			時間	教學資源	教學評量
3-1 參數評估	教學流程	教學情境 vs 學生活動				
	由日常生活之經驗談起	1.在先前的課程中,我們事先給定一組是先調整參數的資料,另外再給一組資料用測試用的,但是在現實生活中我們只有固定一筆資料,測試的資料是事後才能得到的,因此需要對現有的資料進行調整				教學投影片
	教授簡單之預測方法	1.利用現有的資料分成測試用的資料以及調整參數用的資料,透過一部分的資料進行參數的調整,另一部分測試錯誤率為多少 2.在教授背後的方法時,提供一些微調的機制,使出來的結果可以根據使用者的判斷決定				

		該使用何種方法較為恰當			
	利用提供的資料針對上課所教授之方法進行實作	<ol style="list-style-type: none"> 1.提供一組數據來實現上課教授之方法 2.通過比較 MATLAB 內建的函數以及學生自行撰寫之程式進行比較 3.根據參數的調整進行比較，比較在不同參數下之效能 			
3-2 數值最佳化	由日常生活之經驗談起	<ol style="list-style-type: none"> 1.在大部分所遇到的最佳化問題中都是沒有辦法有數學式可以表示最佳解的，因此要利用數值疊代的方法進行求解 2.在解最佳問題時，通常需要在特定範圍裡進行求解，在這邊也會介紹如何在疊代的過程中，把解的範圍一起帶入進行考慮 		教學投影片	
	教授簡單之方法	<ol style="list-style-type: none"> 1.在利用數值求解的過中，要擁有特定形式的最佳化問題才能保證可以得到最佳解，介紹可以有辦法解出最佳解的形式，讓學生在寫功課時能自行判斷是否可以得到最佳解 2.介紹常用的數值疊代方法，並介紹簡單之理論基礎 			
	利用提供的資料針對上課所教授之方法進行實作	<ol style="list-style-type: none"> 1.出多個最佳化問題請同學判斷該問題是否能找到最佳解 2.針對上課所教之數值模擬方法，出一題請學生進行實作，並跟現有之套裝軟體進行比較，觀察自行撰寫之程式跟套裝軟體之差異 			
3-3 叢集 3-4 矩陣分解	由日常生活之經驗談起	<ol style="list-style-type: none"> 1.在之前的分類問題中，我們是針對已分類好的資料進行分類，但是通常我們獲得的資料中要先對資料進行分群，才能進行之前分類的演算法 <ol style="list-style-type: none"> (1)根據現在的網頁中出現之關鍵字，將同一類的新聞放在一起 2.在日常生活中，我們要利用已知的資料來估測未知的資料，因此介紹一些線性代數的理論，來估測未知的數據 <ol style="list-style-type: none"> (1)網頁中的推薦系統，利用現有客戶的評分，推薦新進之客人可能需要的內容 		教學投影片	
	教授簡單之方法	<ol style="list-style-type: none"> 1.介紹一種叢集的演算法，並教授其中之理論，利用現有的函式庫，請同學想一些日常生活中遇到的問題，應用此演算法，並當作期末考試 			

2.在日常生活中，我們獲得的資料，我們可以利用矩陣理論中的奇異質分解 (singular value decomposition, SVD)，利用機器學習的方法把未知的資料猜出來

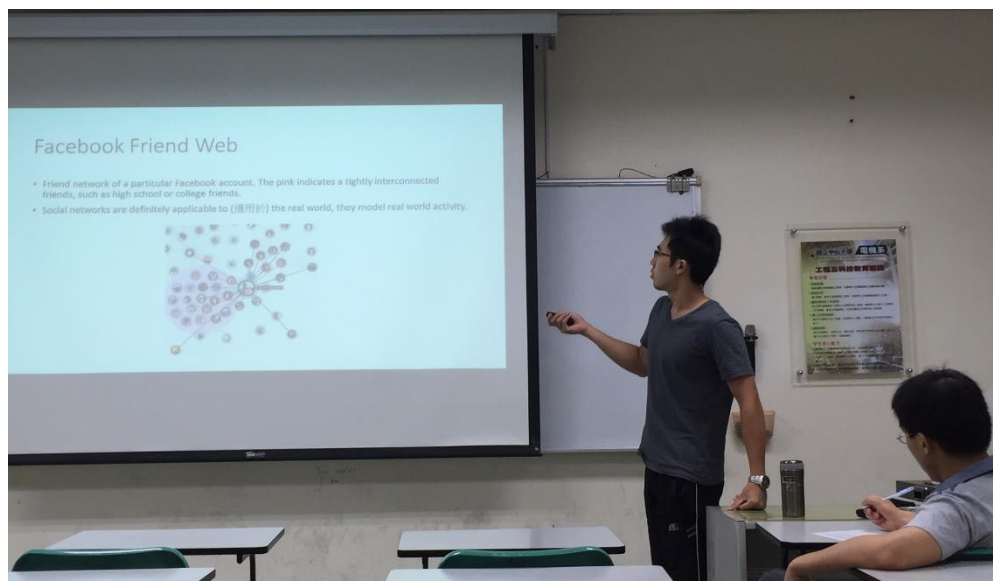
學生成果(活動照片、學生作品、學習單)

利用期末所教之叢集演算法，從日常生活之問題出發，先猜測事件之間的關聯性，並利用網路搜尋或是發問卷的方式，蒐集大量得數據之後，利用叢集演算法將資料分群，利用分類的結果再進階討論同一群之關係，



圖一 六都評比分析

使用前面所學的叢集演算法，藉由許多不同的政府部門的滿意度評比，利用分群出的結果判斷是否與六都的位址有關？



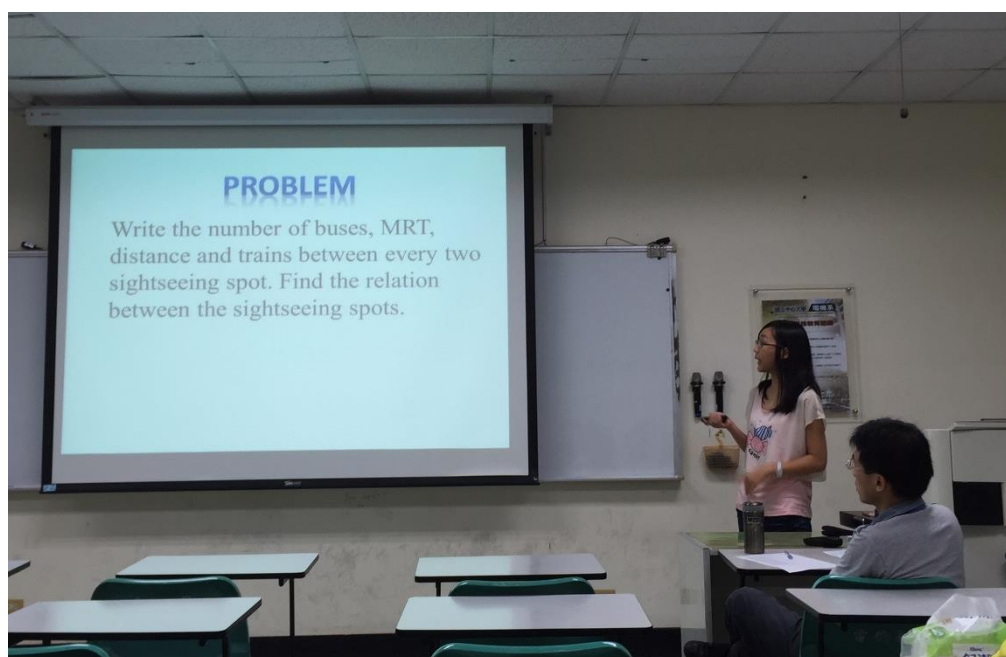
圖二 臉書訊息和朋友的分析

利用臉書的互動資料，例如：點讚、塗鴉牆、動態訊息、…，將這些資料量化做分級評比用來做叢集，探討是否網路上的互動與真實的人際關係是否有關？

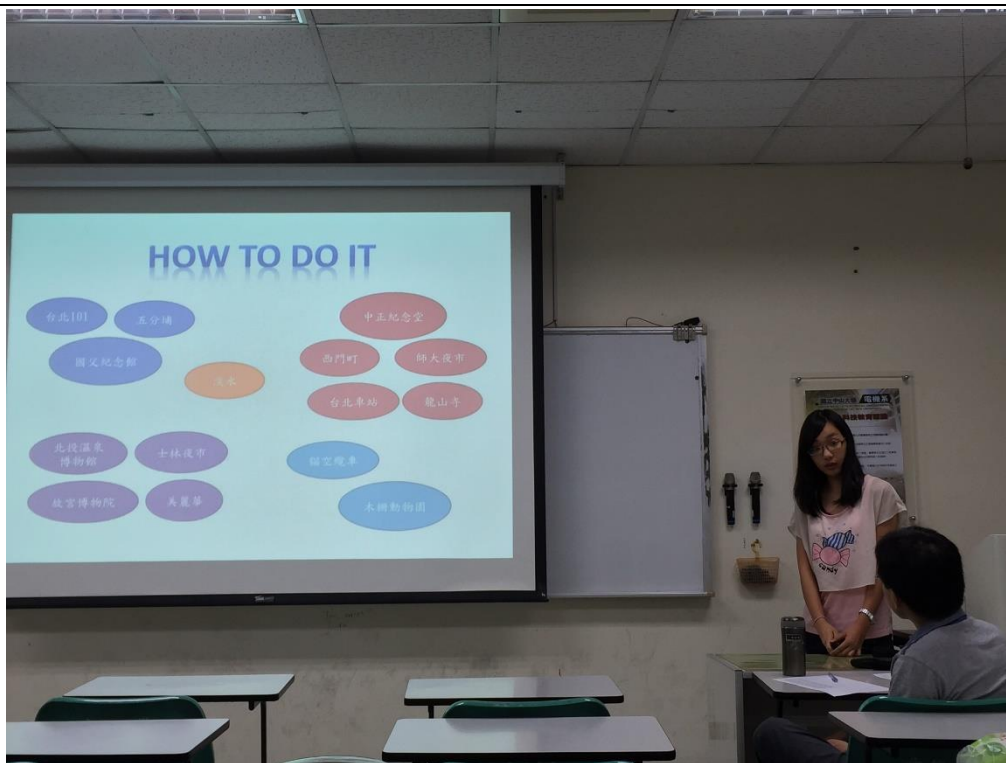


圖三 球拍材質和售價與原產地的關係

利用網路上可取得的球拍資料，如：售價、材質、…，將這些資料分級量化區分再使用叢集分類，探討球拍的價格和材質是否與產地相關，另外透過這些資訊，也可拿來推薦消費者，使消費者能達到滿意的球拍

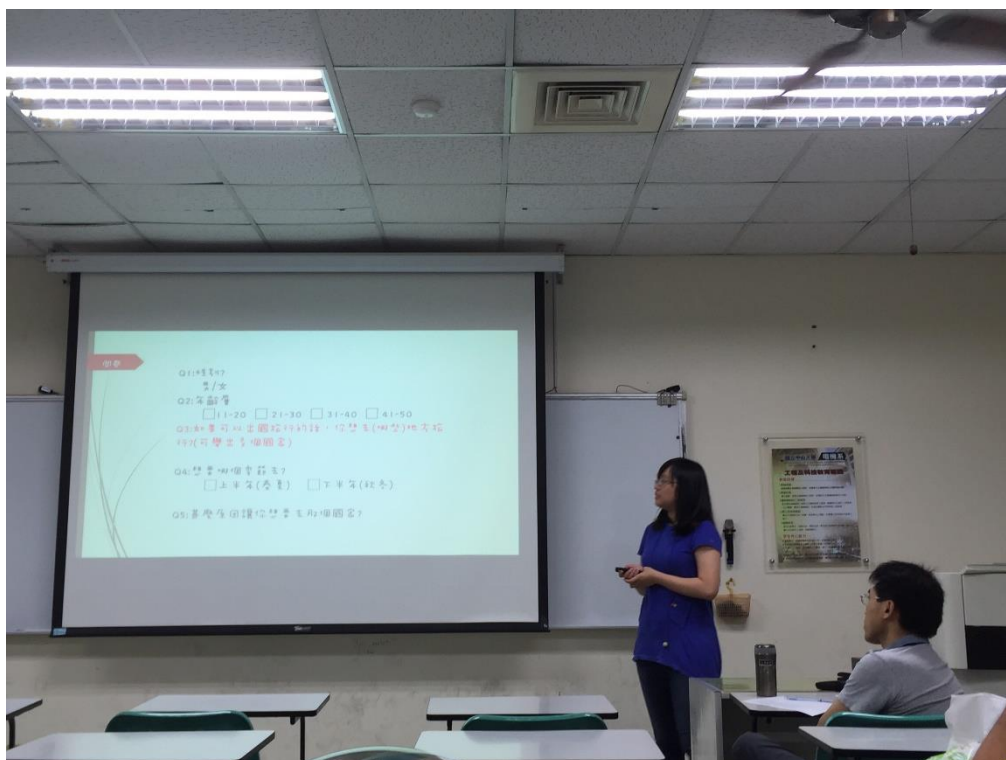


圖四 景點的距離與交通工具的關係



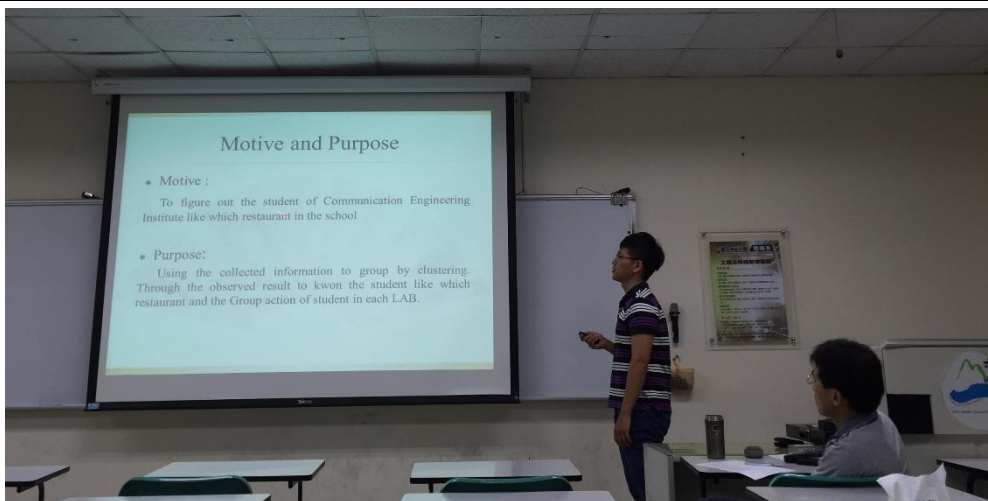
圖五 景點的距離與交通工具分群的結果

將景點的距離經過分類量化和景點的交通工具，經過叢集分類之後，可分出互相方便抵達的景點，作為遊客將來規劃行程時的參考。



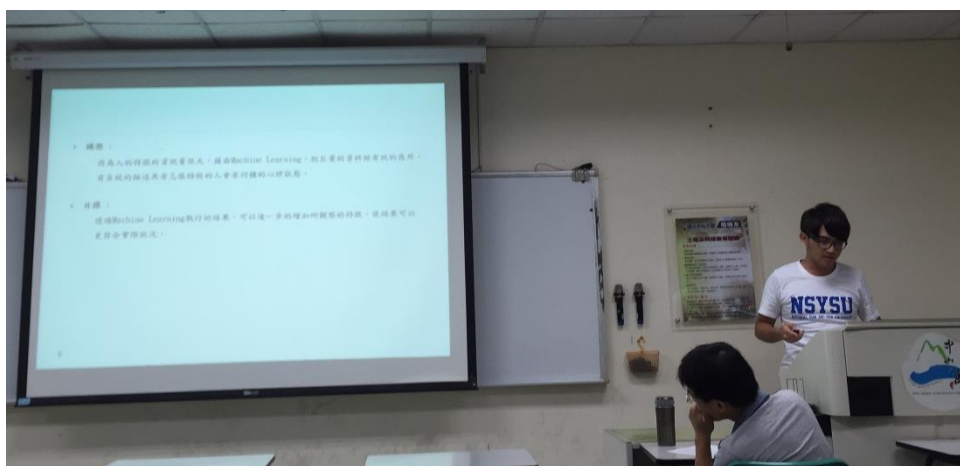
圖六 共同一起旅行的夥伴關係

利用四個因素，年齡、男女、國家、時間，透過問卷的方式，調查不同年齡層跟性別的意願之後，將填問卷的人進行分群，透過分群的方式，可以推薦相同群的人可以一起出遊，增加旅遊的樂趣



圖七 學生一起吃飯與研究所實驗室的關係

利用學生們是否互相一起吃飯為依據，做分群的處理，探討是否一起吃飯是否與實驗室相關



圖八 面相與人格特質的關係

利用網路上可查詢到的圖片，將面相資料做分類，例如：大耳垂、小嘴、...，利用這些資料將所收集的人物做分群的處理，探討是否面相與人格特質(工程、企業、慈善、政治)有關。



圖八 相機配備與使用者的推薦關係

利用相機之性能以及其價格，將同價位中的相機，擁有類似功能的相機推薦給消費者，方便消費者在固定價位中找到最適合之產品



圖九 圖形邊角判斷

利用圖片的幾何圖形，將許多照片依據不同的內容進行非類，例如給許多車跟飛機的照片，將車跟飛機以程式的方式分類出來



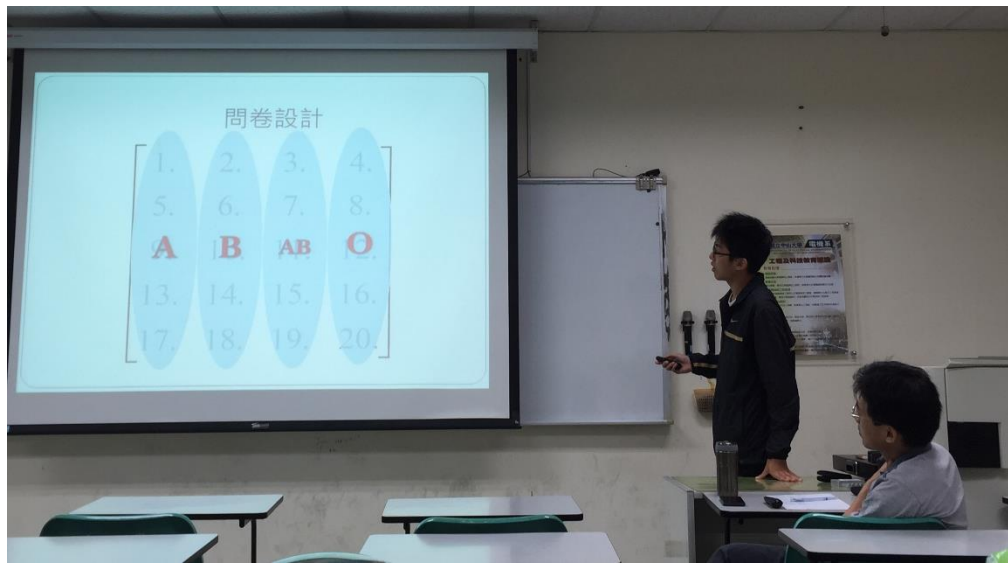
圖九 都市化程度與生活品質關係

以一些客觀的資料將都市分類，例如：人口數、大眾運輸……，將都市進行分類



圖十 準時畢業與實驗室作息的關係

利用一天在實驗室出席之狀況以及研究之狀況，將碩士班的同學進行分類，利用分類的結果找出是否可以同時畢業



圖十一 利用做問卷的結果將人以血型分類

以相同血型會具有類似的思考特徵來設計問卷，並利用問卷的結果將人進行分類，看是否為同一血型