

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

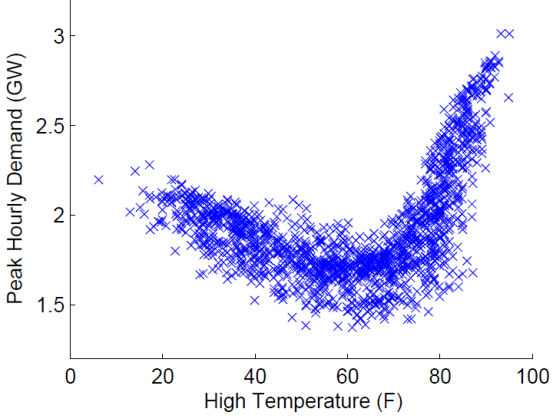
1. 類神經網路(老師授課)
2. 深度學習(老師授課)
3. 逕向基函數網路(老師授課)
4. 矩陣分解(老師授課)
5. 應用上述方法於生活上的應用

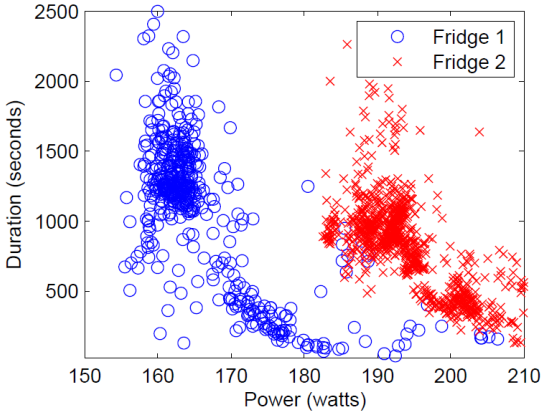
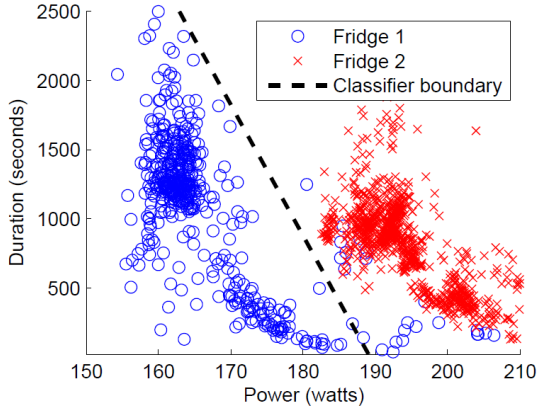
評量

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 1: 迴歸分析與分類系統(基礎) 1. 線性回歸(老師授課) 2. 非線性回歸(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
1-1 線性回歸	教學流程	教學情境 vs 學生活動			
	由日常生活中之經驗談起	1.在日常生活中需要針對許多資料進行預測,首先需要找出資料之相關性 (1)溫度與最高用電量可能會有相關 (2)居住環境之大小與價錢可能會有相關 (3)日期可能跟大眾交通工具乘載量會有相關 (4)太陽能之發電量與當天之氣候可能會有關係 2.在找出欲預測之資料與其他資料的相關性之後,針對上述資料的相關性以 MATLAB 描點先進行觀察,觀察跟一開始猜測的相關性是否符合			
					教學投影片

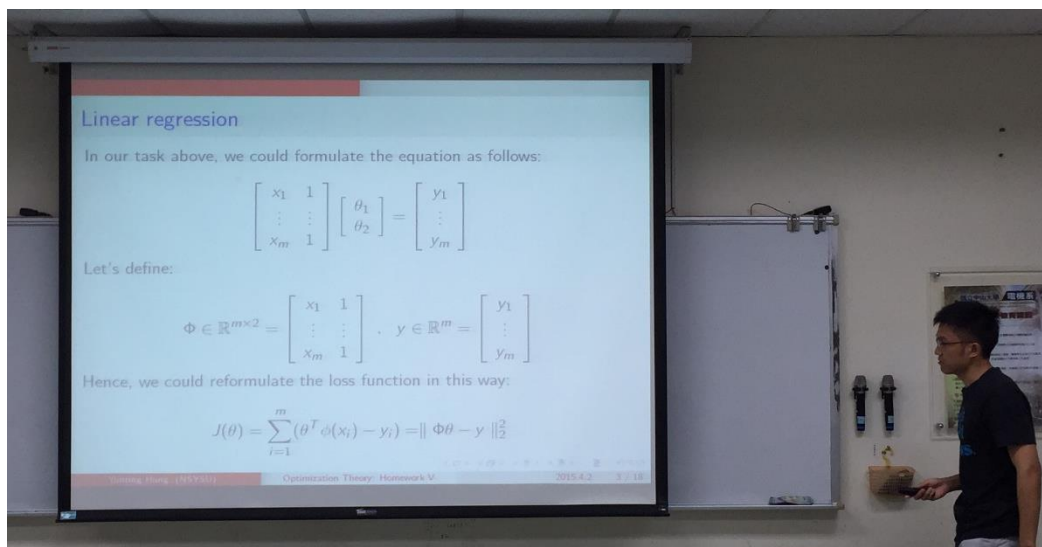
	教授簡單之預測方法	<p>1.在課堂中我們先教授一個簡單的機器學習的方法稱為線性回歸，利用已知的資料來描出一條線近似我們的觀察結果，接著我們以這條線當作我們預測的工具，當我們接受到新的資料時，我們可以藉由這條線來預測可能會出現的結果</p> <p>2.在教授背後的方法時，提供一些微調的機制，使出來的結果可以根據使用者的判斷決定該使用何種方法較為恰當</p>			
	利用提供的資料針對上課所教授之方法進行實作	<p>1.提供一組數據來實現上課教授之方法</p> <p>2.通過比較 MATLAB 內建的函數以及學生自行撰寫之程式進行比較</p> <p>3.根據參數的調整進行比較，比較在不同參數下之效能</p>			
1-2 非線性回歸	以簡單的方法說明本節之重點	<p>1.上一堂課學習的是線性回歸分析，但日常生活中很多的情況是呈現非線性的相關而不是線性的，路如上節課提到的溫度與最高用電量可能會有相關，可以容易得知溫度越高冷氣用電量一定會越高，但是在溫度低的時候，也有可能因為需要開暖氣而使得用電量上升，因此溫度與用電量不一定呈現的是線性關係，所以可以藉由非線性的曲線來預測可能會出現的結果。</p> 		教學投影片	
	教授簡單之預測方法	<p>1.由於非線性回歸問題通常會遇到複雜度及高的計算，課程中教授了一些減少運算複雜上的化簡(kernels)</p> <p>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</p>			

	<p>利用提供的資料針對上課所教授之方法進行實作</p>	<p>3.提供一組數據來實現上課教授之方法 4.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 5.根據參數的調整進行比較，比較在不同參數下之效能</p>			
<p>1-3 分類系統 1-4 分類系統</p>	<p>由日常生活中之經驗談起</p>	<p>1.在日常生活中我們不一定要知道未來的趨勢，而是只要知道是否會超過一個值，只需知道最後結果，不須知道趨勢，以此切入主題</p>			
	<p>教授簡單之分類方法</p>	<p>再以兩種電器的用電量以及用電時間來畫圖，以我們肉眼可以看到可以很明顯可以看出有兩群，以我們肉眼看是非常簡單，但是當有比較極端的值出現時，我們可能也無法區分，因此用程式判斷的話也是非常困難，因此我們需要找出一條線，當之後有新的資料進來時，出現的右邊就可判斷程式紅色那群，反之就是藍色那群</p>  <p>在教授一些簡單的演算法之後，我們可找出成為判斷的一條線，從下圖就可以看出可以非常容易判斷分類的依據</p> 			

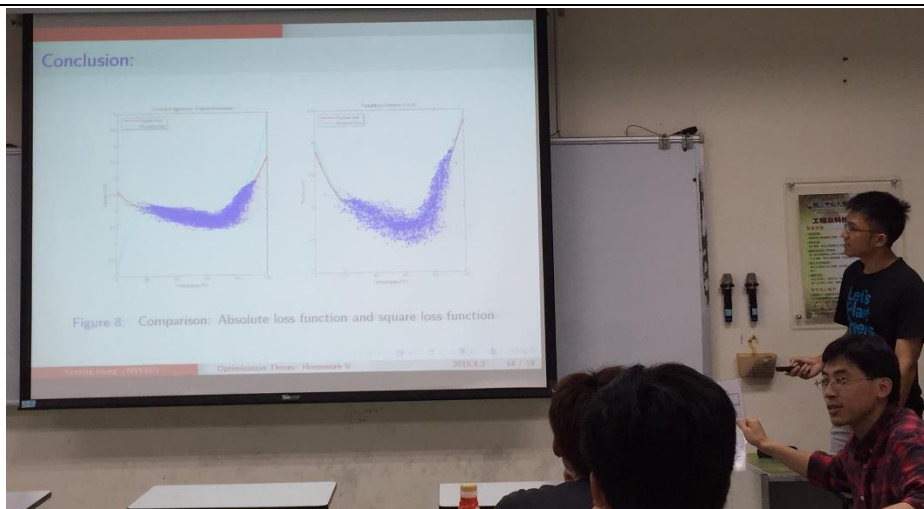
		在課堂中亦提供許多不同之模型供學生參考，並提供模型選擇之依據，供學生在做作業時有不同的選擇並比較不同模型間之差異，並且在執行期末專案時能有更多元的選擇空間			
	利用提供的資料針對上課所教授之方法進行實作	<ol style="list-style-type: none"> 1.先教授垃圾郵件分類的基本概念，利用收集郵件內的關鍵字，接著賦予每個關鍵字不同的權重，接著利用這些權重來判斷這封郵件是否為垃圾郵件 2.在作業中提供學生一些已經擁有的資料來找出最好的權重，利用上課所教之方法，利用找出的權重放入測試資料中，觀察依據不同數目的 train data 數量，查看利用權重判斷是否為垃圾郵件，並計算錯誤率 3.除了上述過程之外也利用不同模型來比較執行時間以及其效能 			
1-5 漫想~ 應用機器學習	由日常生活之經驗談起並請同學結合目前所學觀察日常生活中可運用之地方	<ol style="list-style-type: none"> 1. 利用前面所教授的基礎的迴歸分析讓每位同學各自發想迴歸分析的應用和後續的創意，在課堂上和和其它同學們分享。 2. 分享創意: 捷運的車次間隔、購物推薦系統、疾病預防系統、個人溫控熱水器、個人特質分析器、研發替代役收尋器、紅綠燈最佳變換學習器、棒球收入估測器、電腦性價比估測器、股市預測分析器.....。 			

學生成果(活動照片、學生作品、學習單)

針對第一次作業進行報告，對之前所教授之線性迴歸跟非線性迴歸進行報告，如圖一所示，接著在學生報告完之後，老師針對其餘學生有問題之地方進行較詳細之解說，如圖二所示。

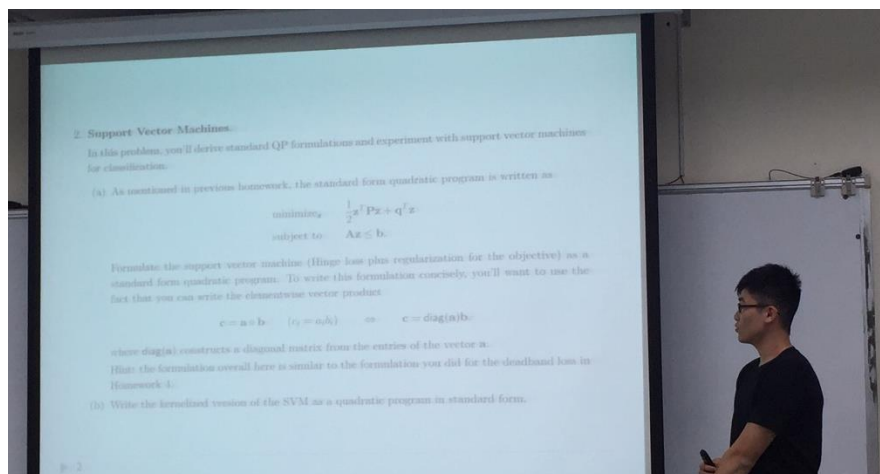


圖一、學生針對作業進行報告

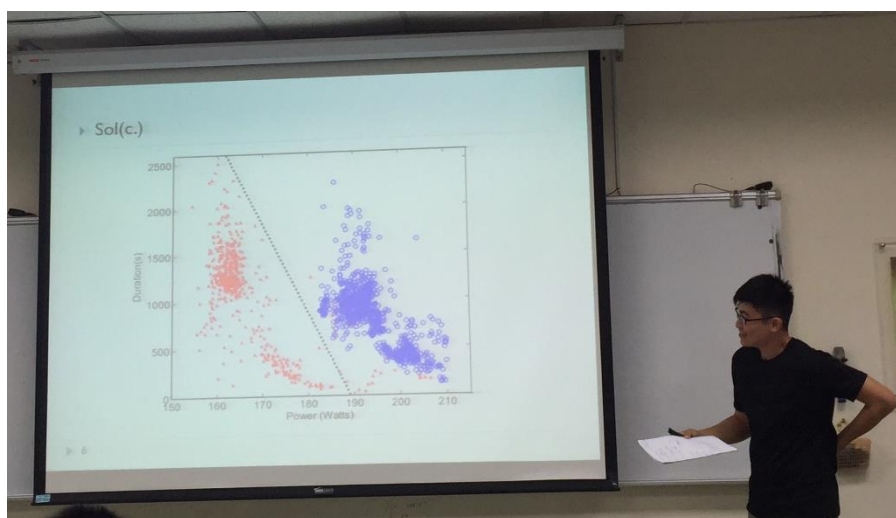


圖二、老師針對有問題的地方進行講解

在教完預測的方法之後，我們也教授分類的觀念以及技巧，接著給同學一組數據，請同學依照上課所教之技巧進行分類以及報告，如圖三以及圖四所示。

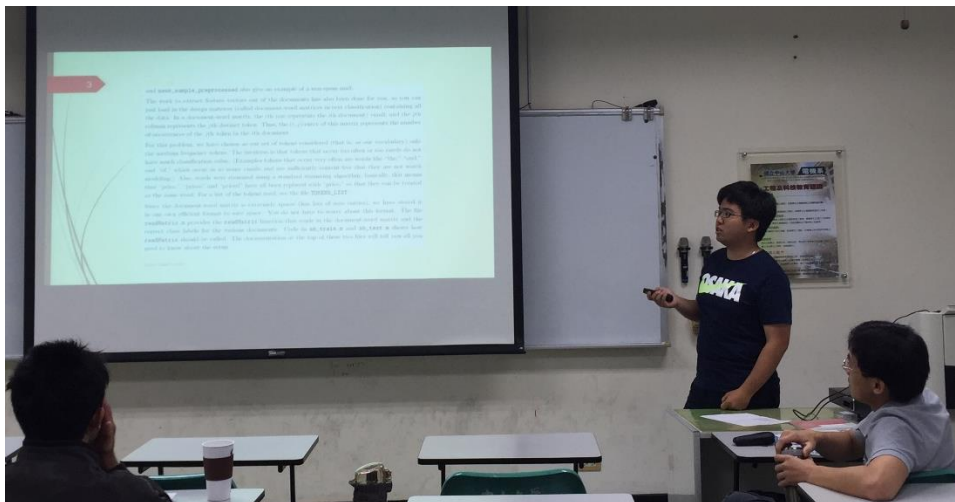


圖三、學生針對作業問題進行描述

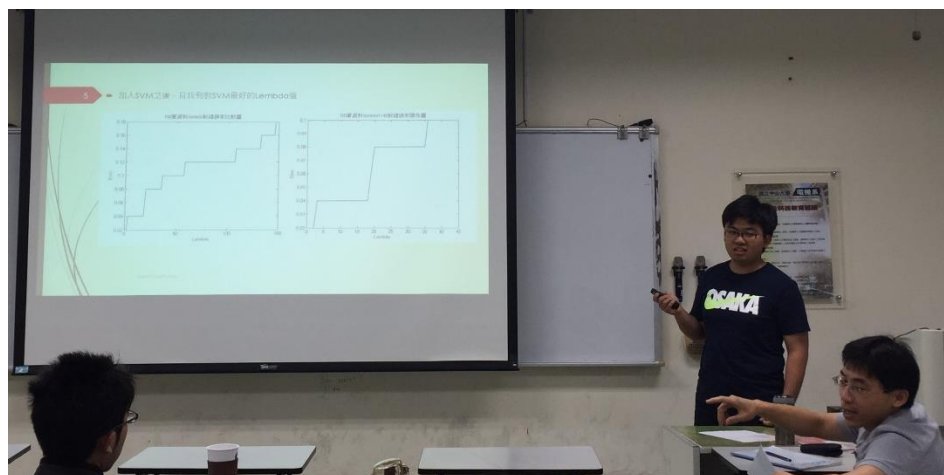


圖四、學生針對作業之結果進行講解

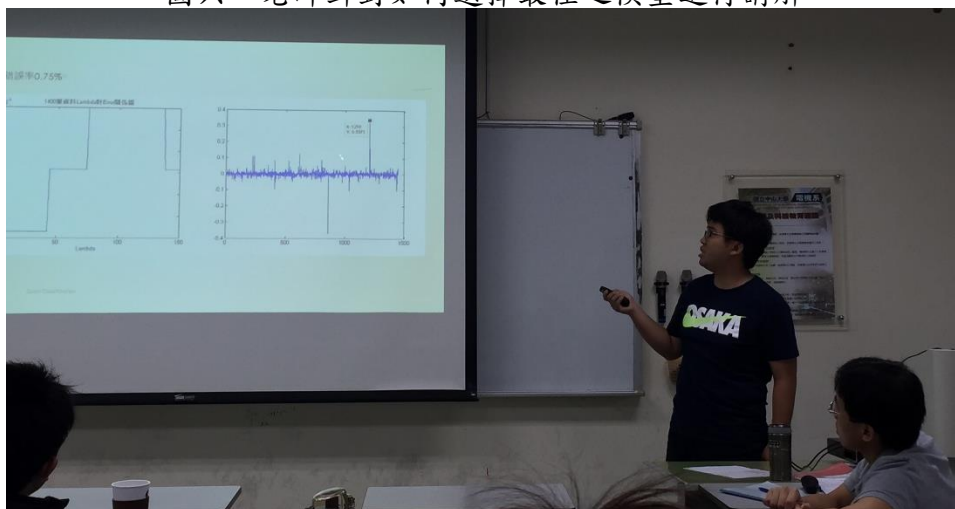
在做完基本的分類方法之後，我們也教授簡單垃圾郵件分類教授分類的觀念以及技巧，接著教同學如何選擇正確的模型在垃圾郵件分類的問題中，接著給同學一組數據，請同學依照上課所教之技巧進行分類以及報告，如圖五、圖六以及圖七所示。



圖五、學生針對垃圾郵件問題之運作原理進行描述



圖六、老師針對如何選擇最佳之模型進行講解

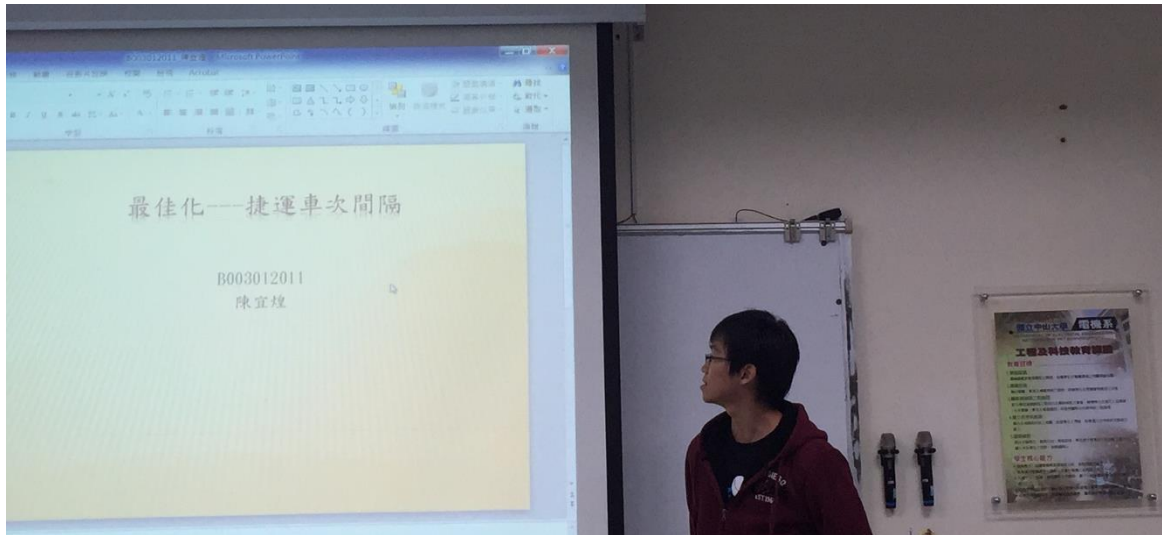


圖七、學生針對結果說明在郵件中出現那些關鍵字是垃圾郵件的機率較高

教學活動省思與檢討

在教授完一些簡單之機器學習之理論之後，請同學先思考目前所學可以運用在那些地方，並請同學進行簡單之報告。

利用機器學習可以根據人的流量以及一些其他因素分配捷運車次間隔，期望可以對人群之疏散可以有更好的效果。



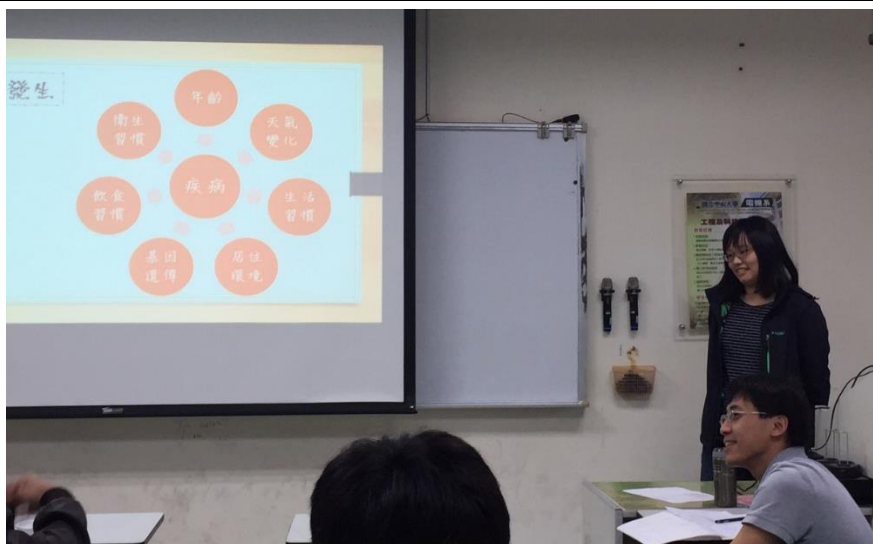
圖八、捷運車次間隔之預測

在線上購物中，用機器學習可以根據人的流量以及一些其他因素分配捷運車次間隔，期望可以對人群之疏散可以有更好的效果。



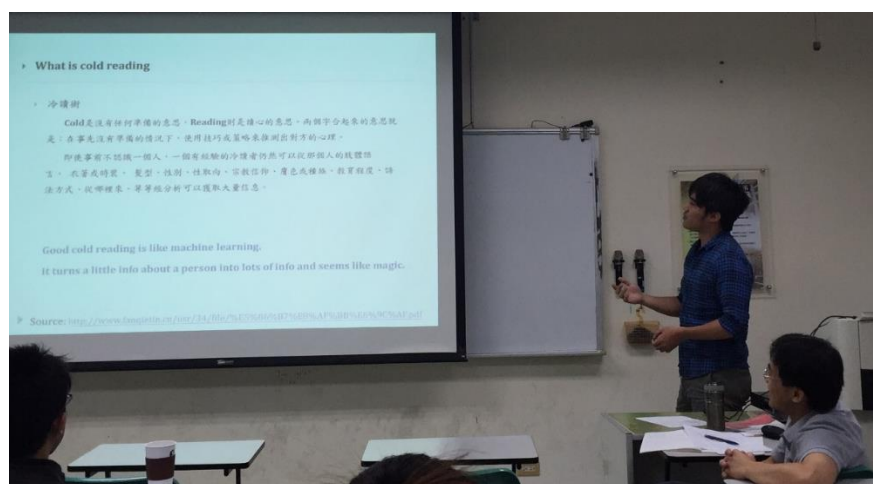
圖九、線上購物之推薦系統

學生在看完電影「大英雄天團」中獲得靈感，提出可以藉由一些生理上以及其他之因素，預測疾病發生之可能性以及時間點，



圖十、疾病之預測

學生提出藉由觀察一個人的特質可以將人進行分類，另外學生也提出藉由所學之專長讓公司可以藉由分類使得公司可以找到最符合公司要求的學生。

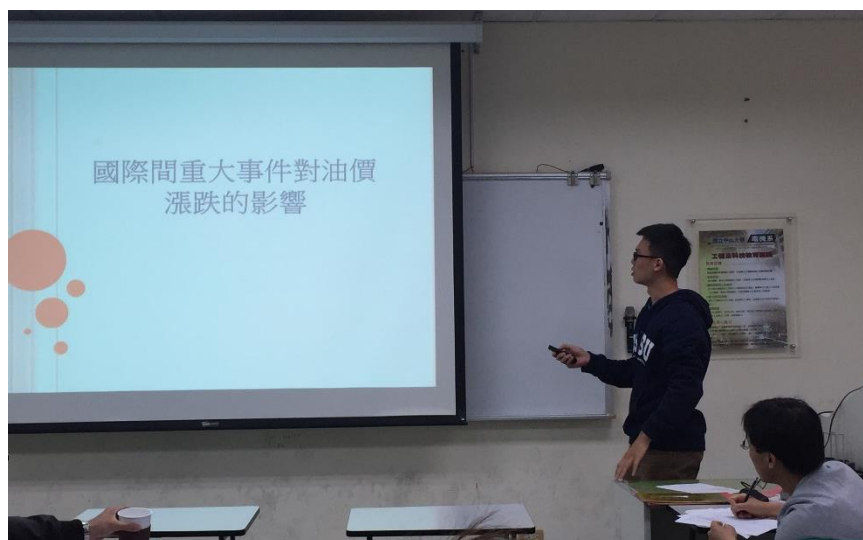


圖十一、藉由觀察一個性質判斷一個人的性格



圖十二、研發替代役之推薦系統

學生提出每次國際油價有所變動時，可以藉由前面學到之線性以及非線性回歸來預測當變動幅度到達多少時，對可能發生之事件進行預測。

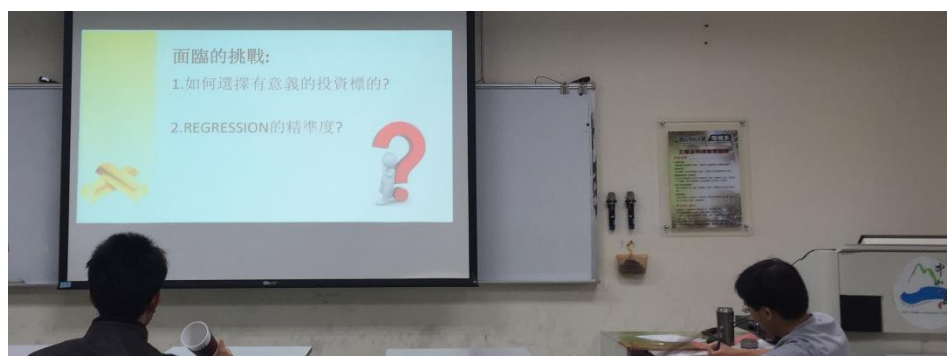


圖十三、國際油價變動影響之預測

學生提出針對每個公司所發表的產品以及財報等等因素，對該公司之股價進行預測，使得投資能夠有最小之風險。



圖十四、針對公司發生的事件預測其股價之變化



圖十五、要預測股價所需要的一些分析