行銷資料科學課後心得

B10409047 王彥琳

此次上課，老師教導我們如何找出有用的變數，以及說明建設模型後，只要透過機器模型自動處理，就可以用不到一秒鐘的時間來找出變數。

而提到變數，究竟甚麼是變數?還有如何判斷特徵變數是正向還是負向?

這邊有舉一實例:向范冰冰打扮穿著艷麗的人，推銷動漫產品，較不會購買產品，因此可推斷出，穿著豔麗的人可能較不會購買產品的比例較一般人高。

而機器學習等於找一個最好的函數，產出結果，其分別是透過以下兩種辨識。

1. 人類辨識
2. 對話系統

資料應用白話來說，就是從顧客口袋中掏錢，其可採用的方式如下:

1. 精準行銷
2. 個人化寫手

而此次課程裡。老師也有講解隨機森林(RF)、強大演算法、運用群體智慧、模型實戰等方式的概念化想法。

由於對隨機森林的好奇，變上網搜尋關於隨機森林的深入資料，其解釋隨機森林是一種多功能的機器學習算法，能夠執行回歸和分類的任務。同時，它也是一種數據降維手段，用於處理缺失值、異常值以及其他數據探索中的重要步驟，並取得了不錯的成效。另外，它還擔任了集成學習中的重要方法，在將幾個低效模型整合為一個高效模型時大顯身手。

在隨機森林中，我們將生成很多的決策樹，並不像在CART模型里一樣只生成唯一的樹。當在基於某些屬性對一個新的對象進行分類判別時，隨機森林中的每一棵樹都會給出自己的分類選擇，並由此進行「投票」，森林整體的輸出結果將會是票數最多的分類選項；而在回歸問題中，隨機森林的輸出將會是所有決策樹輸出的平均值。