

作業研究(二)

授課老師:郭清章

研究室:中商大樓5F (應用統計系)(7517)

Office hour:(五)16:10-18:00

Email:johnkuo@nutc.edu.tw

人工智慧演算法

- 啟發式演算法

相對於最優化演算法（求得該問題每個例項的最優解）提出的。

- 啟發式演算法定義：

一個基於直觀或經驗構造的演算法，在可接受的花費（指計算時間和空間）下給出待解決組合優化問題每一個例項的一個可行解，該可行解與最優解的偏離程度一般不能被預計。

啟發式演算法

- 通俗的說：

咱們都生活在某個小區，你看到你鄰居每天晚上擺地攤賺錢，你受到啟發也每天去同樣的地方擺地攤賺錢，然後越來越多的人發現擺地攤賺錢，就都出來擺地攤了，就這麼個道理。

仿生演算法

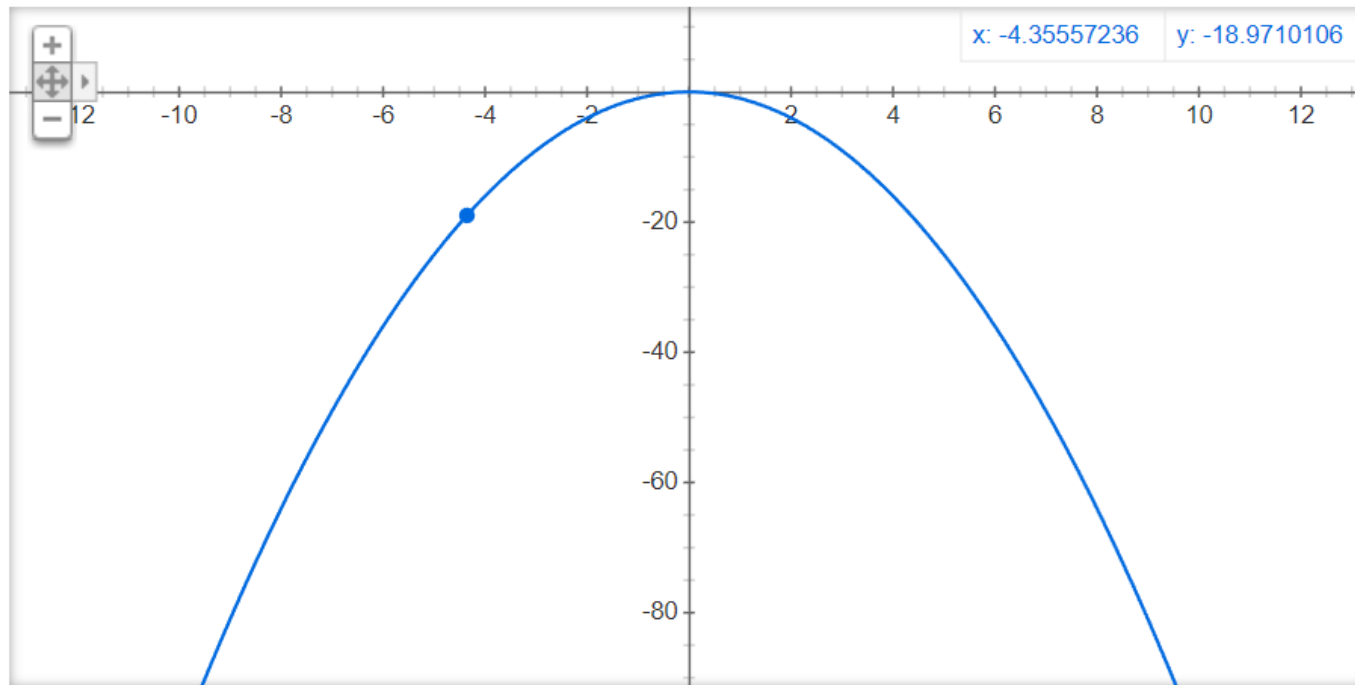
- 仿動物類的演算法：粒子群優化，螞蟻優化，魚群演算法，蜂群演算法等；
- 仿植物類的演算法：向光性演算法，雜草優化演算法，等等；
- 仿人類的演算法有：和聲搜尋演算法是較好的演算法；

單粒子演算法

- 區域搜尋法 (Local Search)
- 貪婪算法 (Greedy Algorithm)
- 爬山算法 (Hill Climbing)
- 隨機式爬山算法 (Stochastic Hill Climbing)
- 最陡坡降法 (Gradient Descent)
- 隨機坡降法 (Stochastic Gradient Descent)
- 大洪水演算法 (Great Deluge Algorithm)
- 禁忌搜尋法 (Tabu Search)
- 模擬退火法 (Simulated Annealing)
- 量子退火法 (Quantum Annealing)
- 隨機穿隧法 (Stochastic Tunneling)
- 反傳遞演算法 (Back Propagation)
- 隨機散播搜尋法 (Stochastic Diffusion Search)
- 牛頓法 (Newton's Method)
- Delta 法則 (Delta Rule)

爬山演算法 (Hill Climbing)

$-(x^2)$ 的圖表



多粒子演算法

- [遺傳演算法](#)（GA）支援一系列的解決方案。解的合併或突變增加了解集，其中品質低劣的解被丟棄，尋找進階解決方案的過程模仿了這一演變。
- [類比退火](#)（SA）是一個全域最佳化相關的通過產生目前解的相鄰解來遍歷搜尋空間的技術。進階的相鄰解總是可接受的。低階的相鄰解可能會根據基於品質和溫度參數差異的概率被接受。溫度參數隨著演算法的行程被修改以改變搜尋的性質。
- [反作用搜尋](#)最佳化的重點在於將機器學習與最佳化的結合，加入內部回饋迴路以根據問題、根據實體、根據目前解的附近情況的特點自動調整演算法的自由參數。
- [禁忌搜尋](#)（TS）類似於類比退火，他們都是通過測試獨立解的突變來遍歷解空間的。而類比退火演算法對於一個獨立解只生成一個突變，禁忌搜尋會產生許多變異解並且移動到產生的解中的符合度最低的一個。為了防止迴圈並且促進在解空間中的更大進展，由部分或完整的解組建維繫了一個禁忌列表。移動到元素包含于禁忌列表的解是禁止，禁忌列表隨著解遍歷解空間的過程而不斷更新。
- [人工免疫系統](#)（AIS）演算法仿照了脊椎動物的免疫系統。
- [粒子群最佳化](#)（PSO），群智慧型方法。
- [引力搜尋演算法](#)（GSA），群智慧型方法。
- [蟻群群集方法](#)（ACCM中），這個方法利用了群集方法擴充了蟻群最佳化。
- [隨機傳播搜尋](#)（SDS），基於代理的概率全域搜尋和最佳化技術，最適合於將目標函式分解成多個獨立的分布函式的最佳化問題。

粒子群演算法(鳥群演算法)PSO

「粒子群演算法」採用哪些鳥類覓食的觀念?

「粒子群演算法」主要採用了下列兩個觀念

依照自身經驗決定飛行方向

參考別人的經驗決定飛行方向

這兩件事就像人類的學習過程

人類在處理事情時，通常都會依照自己的方法

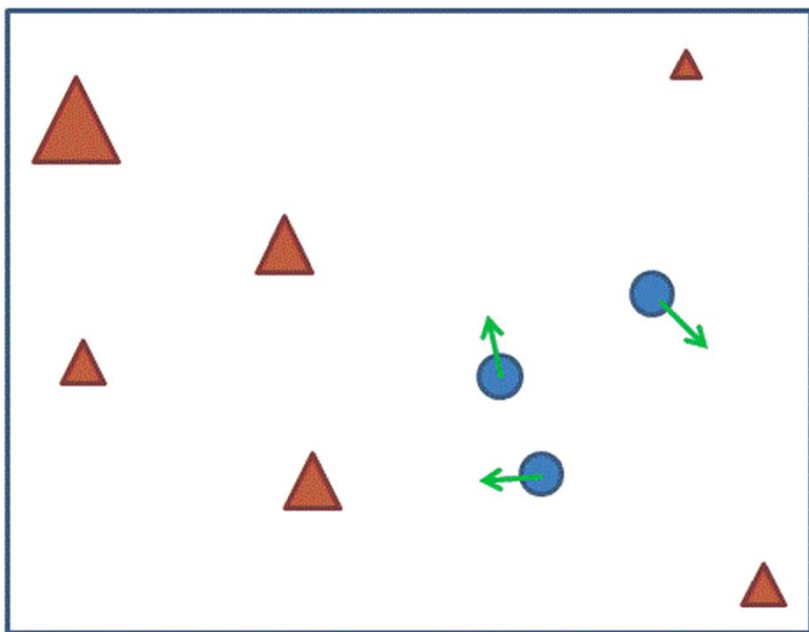
當我們發現有人的方法比自己的方法更好，就會參考並學習該方法

鳥類覓食

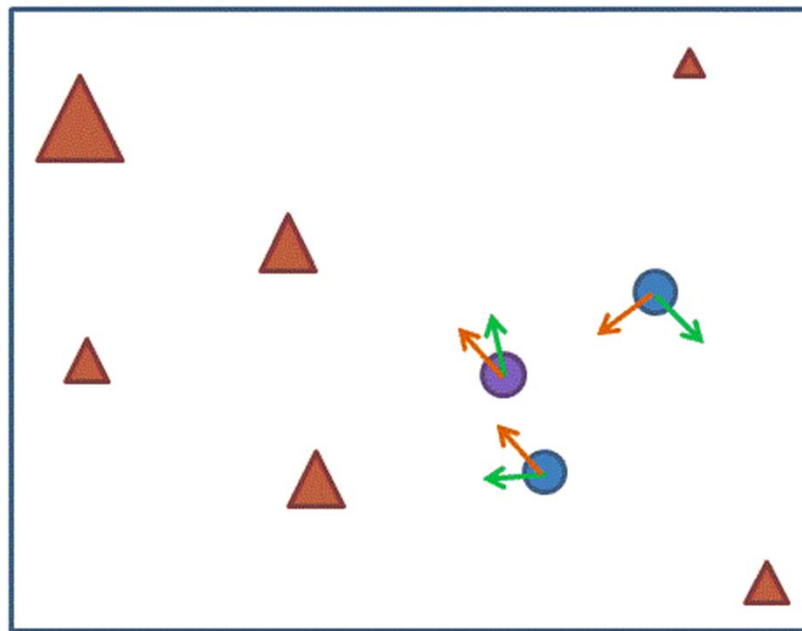


粒子是依照「發現別人比自己更好，就會參考並學習」

● 粒子 ▲ 食物(越大代表越多食物)
→ 粒子自身方向

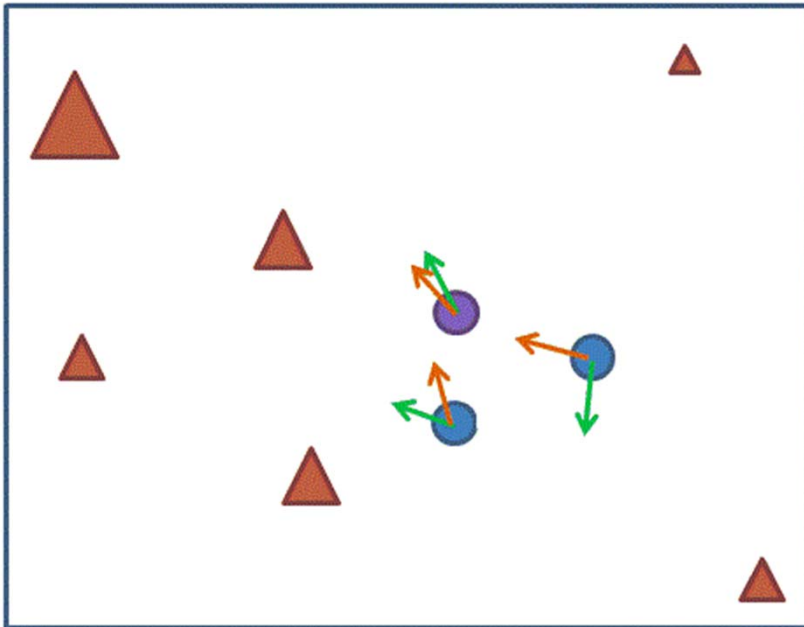


● 最佳粒子
→ 粒子自身方向 → 粒子參考群體方向

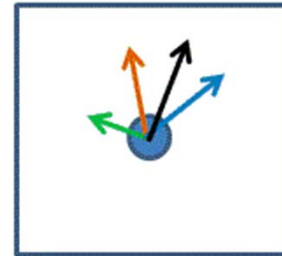


「粒子」的移動方式?

- 最佳粒子
- 粒子自身方向
- 粒子參考群體方向



- 粒子上個時間的方向 (速度)
- 群體最佳方向
- 粒子目前時間的總合方向
- 自身最佳方向



想成為資料科學家？來挑戰Google、FB、Apple等六間公司人工智慧最新面試題

- <https://www.inside.com.tw/article/12649-data-scientist-interview>

微軟

- 三個朋友告訴你今天會下雨，他們中每個人都有三分之一的機率說謊，那麼今天下雨的可能性是多少？
- 如何從產品使用者群中隨機選擇一個樣本？
- 你如何找到百分位數？請編一個程式。

Amazon

- 邏輯回歸模型中如何知道係數是什麼？
- 給出一個長條圖並且想像你正從上面倒水，如何確定長條圖中可以保存多少水？
- 你如何在天平上重複秤重 9 個彈珠 3 次以選擇最重的彈珠？
- 查找西雅圖客戶過去 6 個月中前 10 名利潤最高的產品的累計總和。
- 邏輯回歸和線性回歸的假設是什麼？
- 項目位置 A 的項目機率為 0.6，項目位置 B 的機率為 0.8。在 Amazon 網站上找到該項目的機率是多少？
- K-means 如何運作？你會選擇什麼樣的距離度量？如果不同的特徵有不同的動態範圍呢？

Facebook

- 從 100 個硬幣中隨機抽取一枚硬幣：1 枚不公平的硬幣（都是正面），99 枚公平的硬幣（一正一反）並投擲 10 次。如果結果是 10 個正面，那麼這枚硬幣是不公平的機率為多少？
- 在 Python 中為數值編寫排序演算法。
- 你即將坐飛機去西雅圖。你想知道你是否應該帶一把雨傘。你可以給 3 個隨機的朋友打電話，每個人都會獨立詢問是否下雨。你的每個朋友都有三分之二的機會告訴你真相，並有三分之一的機會對你撒謊。所有 3 位朋友都告訴你正在下雨。在西雅圖實際下雨的可能性有多大？

Facebook

- 考慮一個有 2 名玩家 A 和 B 的比賽。A 有 8 個棋子，B 有 6 個棋子。比賽進行如下。首先，A 滾動一個公平的六面體，並且六面體上的數字決定 A 從 B 接收多少個寶石。接下來，B 滾動相同的六面體，並且完全相同的事情發生在相反的位置。本輪結束。誰在比賽結束時擁有更多的寶石則贏得比賽。如果玩家在回合結束時獲得相同數量的寶石，則會形成平局並且接下來會有一輪。B 在 1, 2, ..., n 輪獲勝的機率是多少？
- 製作 2 個變量的直方圖。
- 你在一個賭場擲骰子，如果擲出 5 則贏，並獲得 10 美元的獎金。你能賺多少？如果你一直玩到你贏了（不管花多長時間），那麼你的預期支出是多少？

Google

- 什麼是 $1/x$ 的導數？
- 繪製曲線 $\log(x + 10)$
- 如何設計客戶滿意度調查？
- 擲硬幣十次，得到結果為 8 個正面和 2 個反面。如何分析一枚硬幣是否公平？
p 值是什麼？
- 你有 10 個硬幣。你每擲硬幣 10 次（共 100 次）並觀察結果。你會修改你的方法來測試硬幣的公平性嗎？
- 解釋一個不是正態的機率分佈以及如何應用它？

Google

- 為什麼使用特徵選擇？如果兩個預測因子高度相關，那麼對邏輯回歸中的係數有什麼影響？係數的信賴區間是多少？
- 如果回歸模型中的兩個係數估計值中的每一個都具有統計顯著性，那麼您是否期望兩者的測試仍然很重要？

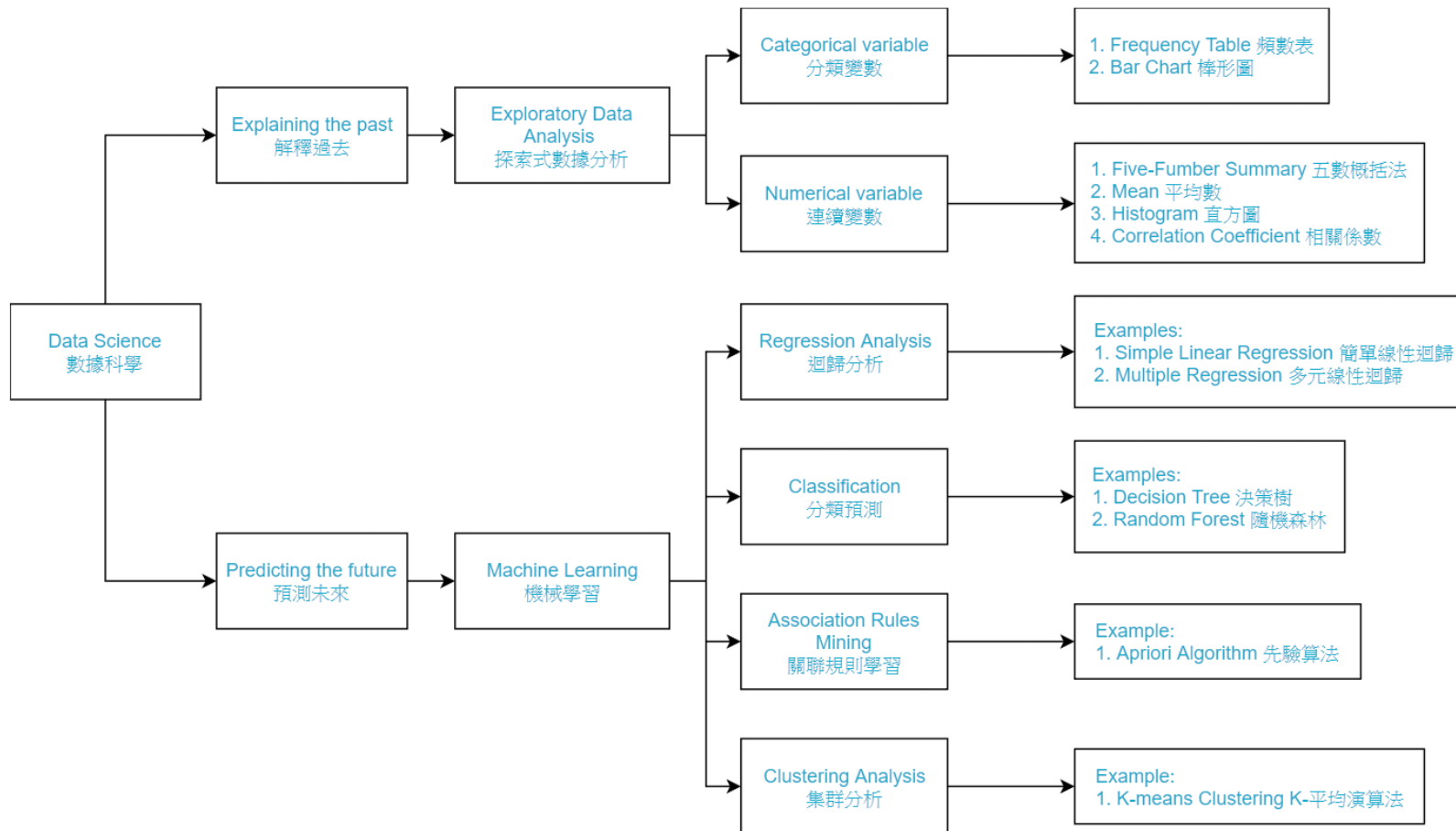
Uber

- 使用隨機白努利試驗來編寫函數以返回來自常態分布的數值樣本
- P 值是什麼意思？
- 解釋線性回歸、線性假設和線性方程
- 解釋 Logistic 回歸、Logistic 假設和 Logistic 方程

Part IV : 各週授課進度與內容

週次	主題內容/章節	備註
第一週	簡介	
第二週	基礎演算法	Flow chart
第三週	基礎演算法	演算法
第四週	Matlab程式語法	
第五週	Matlab程式語法	放假
第六週	啟發式演算法1	爬山演算法
第七週	啟發式演算法2	梯度下降法
第八週	啟發式演算法3	貪婪算法
第九週	繳交期中報告	Propose(提案一頁)
第十週	啟發式演算法4	模擬退火算法
第十一週	啟發式演算法5	GA(基因演算法)
第十二週	啟發式演算法6	螞蟻演算法
第十三週	啟發式演算法7	PSO(粒子演算法)
第十四週	啟發式演算法8	回歸
第十五週	啟發式演算法9	決策樹
第十六週	啟發式演算法10	K-means(k-平均演算法)
第十七週	期末報告	
第十八週	期末報告	

數據科學入門 – 思維和工具



- **數據科學 (Data Science)** 是通過編程來分析和運用數據以解釋過去和預測未來的學問。數據科學是一個多學科技能組合的領域，它結合了**統計學 (Statistics)**，**機器學習 (Machine Learning)** 和**數據庫 (Database)** 技術。在大數據急速發展的時代，數據科學在未來生活中的應用價值變得越來越高。企業和組織在多年的運營中儲存了大量數據，用數據科學的方法能夠從這些數據中提取非常有價值的知識，加上機器學習和人工智能的發展加快普及。企業和組織能將提取的知識應用到更廣更深層次的場景中。

-

數據科學的工作方向 – 解釋過去及預測未來

- 如要用一句話來說明數據科學家的工作性質是什麼的話，「解釋過去及預測未來」就是最直觀的了解。數據學家在解釋過去和預測未來這兩個方向上有兩套完全不同的方法論和工具，所以開始一個項目時一定要界定清楚問題的本質和選擇正確的工具。
- 解釋過去的數據科學方法：[探索式數據分析 \(Exploratory Data Analysis\)](#)
- 預測未來的數據科學方法：[機器學習 \(Machine Learning\)](#)

- **探索式數據分析 Exploratory Data Analysis**
- 探索式數據分析是通過統計學的方法和可視化技術來描述數據，找出數據的重點以進行進一步分析或歸因。在進行探索式數據分析時，最重要的工作之一是把數據定性為：
 - 分類變數 Categorical variable
 - 連續變數 Numerical variable
- 由於兩種數據所用的分析工具是完全不同的，在編程時把數據定性是最重要的步驟。

- **分類變數 Categorical variable**

- 分類變數是具有兩個或更多類別的變數，如男女、低中高、冷暖熱等變量。分析工具有頻數表和棒形圖等。

- **連續變數 Numerical variable**

- 連續變數可直接了解為可量化數值，是有大小之分、連續不斷的變量（例如，身高，體重，溫度，血糖，.....）。分析工具有五數概括法、平均數、直方圖和相關係數等。

機器學習 Machine Learning

- 機器學習是透過創建數據模型和演算法，利用電腦的高運算能力來預測結果的過程。根據不同的工作目的、數據和事件類型，可選擇的模型和演算法組合可以是千變萬化。加上每日都有數據科學家推出新的演算法，初學者在學習時或會感到無從入手。筆者認入門級的同学最重要是先掌握一些比較常用的演算法類別，再根據遇上的個案慢慢進深學習。
- 根據筆者的經驗有 4 類基本演算法需要先學會：
- [迴歸分析 Regression Analysis](#)
- [分類預測 Classification](#)
- [關聯規則學習 Association Rules Mining](#)
- [集群分析 Clustering Analysis](#)

- **迴歸分析 Regression Analysis**

- 迴歸分析是通過利用一個或多個變數（可以是連續或分類變數）來構建模型，預測目標變數的**數值**（如股價、人數）。

可用工具如：

- 簡單線性迴歸 Simple Linear Regression Analysis
- 多元線性迴歸 Multiple Regression

- **分類預測 Classification**

- 分類預測是通過利用一個或多個變數（可以是連續或分類變數）構建模型，預測樣本所屬的**類別**（如預測一個銀行客戶會否違約，這裡**會違約**和**不會違約**就是兩個類別）。

可用工具如：

- 決策樹 Decision Tree
- 隨機森林 Random Forest

- **關聯規則學習 Association Rules Mining**

- 關聯規則學習的目標是在數據庫中尋找有趣關係和發掘知識，經常被應用於零售業，讓公司研究哪些商品會同時被顧客購買。
可用工具如：

- 先驗算法 Apriori Algorithm

- **集群分析 Clustering Analysis**

- 集群分析的目標是把特性相似的樣本聚集再分成群組。
可用工具如：

- K-平均演算法 K-means Clustering

評分項目

- 出席率及學習態度：40%
 - 報告：60%
1. 期中提出期末研究主題:任一啟發演算法(一頁)
 2. 期末報告上台報告5-10分鐘內完成
 3. 再將期末報告上傳(以ppt上傳)

Q&A

參考資料:<https://dotblogs.com.tw/dragon229/2013/01/10/87127>

<http://ccckmit.wikidot.com/so:singleparticle>

<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/543984/>

<http://skycatcheredu.com/%E6%95%B8%E6%93%9A%E7%A7%91%E5%AD%B8%E5%85%A5%E9%96%80-%E6%80%9D%E7%B6%AD%E5%92%8C%E5%B7%A5%E5%85%B7/>

以上資料僅供教學使用