

Day 9

資料清理數據前處理

相關係數簡介



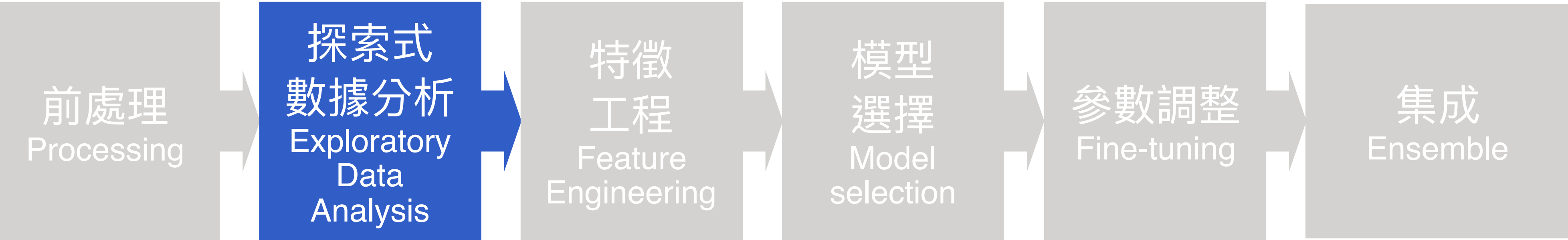
出題教練

游為翔 / 杜靖愷

知識地圖 探索式數據分析 相關係數簡介

機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning

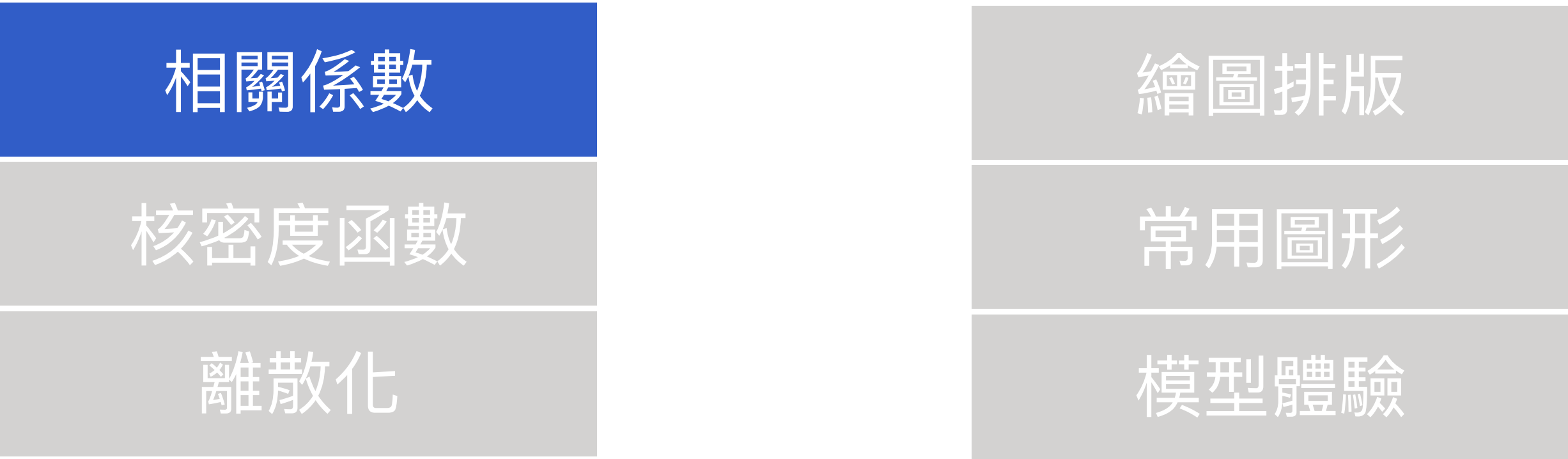


非監督式學習 Unsupervised Learning



探索式數據分析 Exploratory Data Analysis (EDA)

統計值的視覺化



本日知識點目標

了解相關係數

Correlation Coefficient

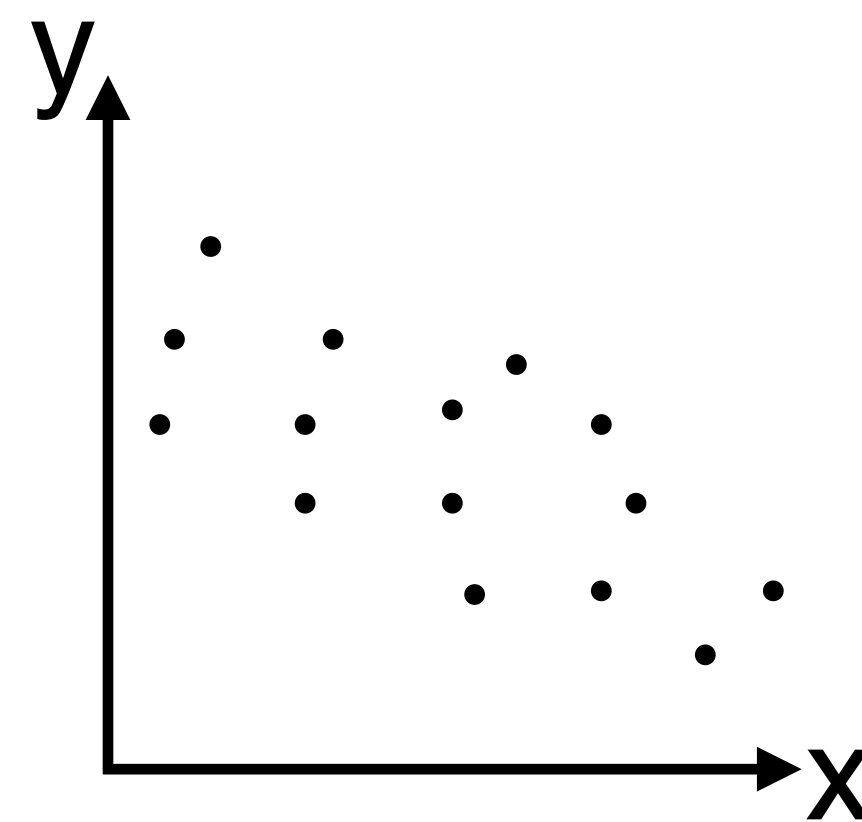
相關係數是其中一個常用來了解各個欄位與我們想要預測的目標之間的關係的指標。相關係數衡量兩個隨機變量之間線性關係的強度和方向。雖然不是表示變數之間關係的最好方法，但可以提供我們很直觀的了解。

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

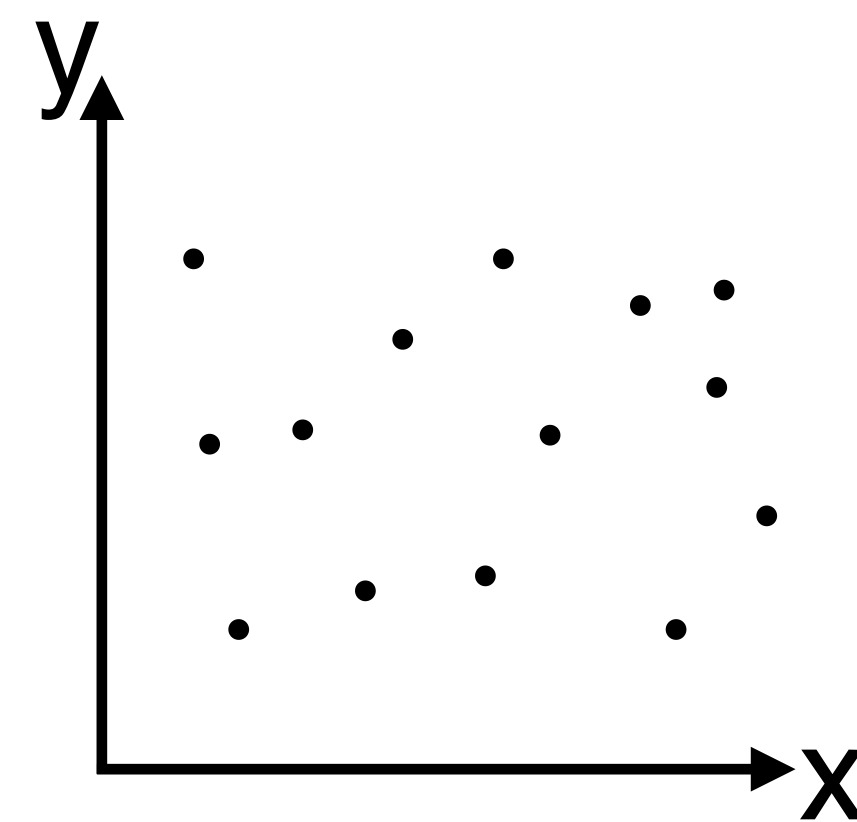
Correlation Coefficient

相關係數是一個介於 $-1 \sim 1$ 之間的值，負值代表負相關，正值代表正相關，數值的大小代表相關性的強度

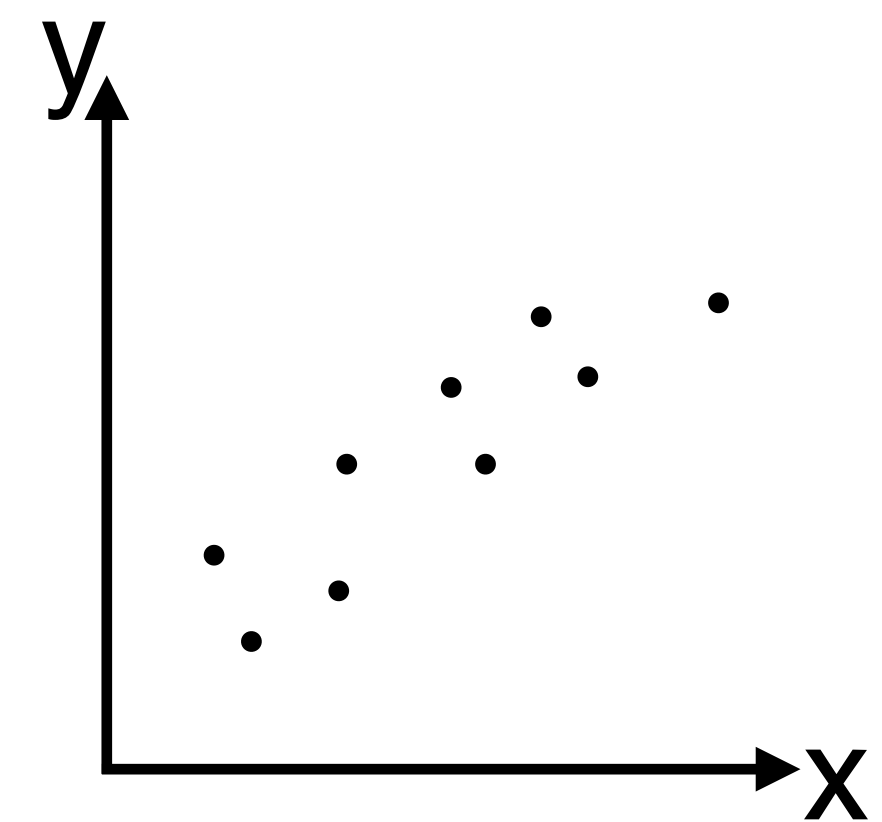
- .00-.19: 非常弱相關
- .20-.39: 弱相關
- .40-.59: 中度相關
- .60-.79: 強相關
- .80-1.0: 非常強相關



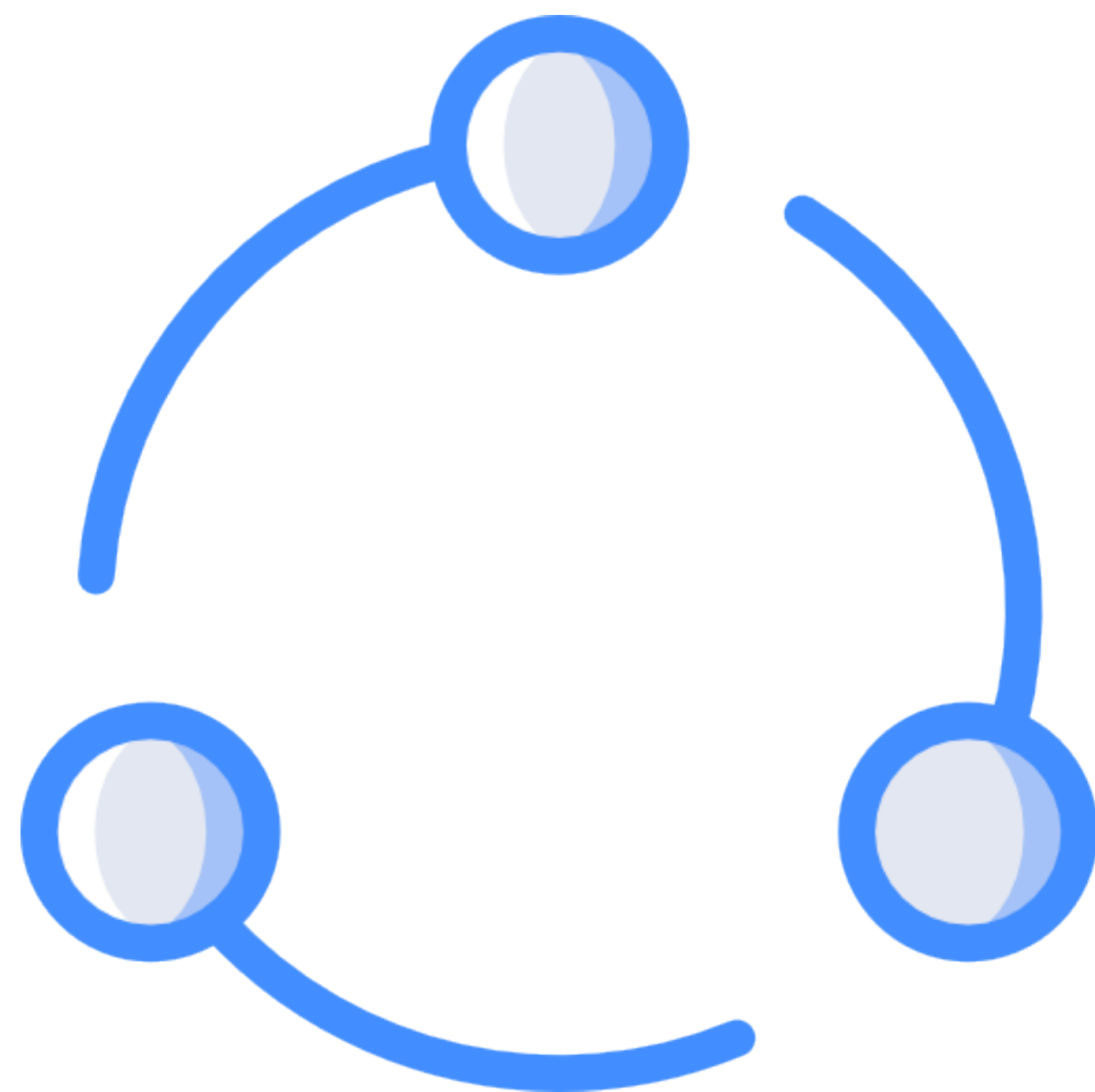
Negative
Correlation



No
Correlation



Positive
Correlation



基本的統計數值只能了解一個變數，如果我們想要了解兩個變數之間的**線性關係**時，相關係數是一個還不錯的簡單方法，能給出一個 $-1 \sim 1$ 之間的值來量化兩個變數之間的關係。

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

