

Day 31

機器學習

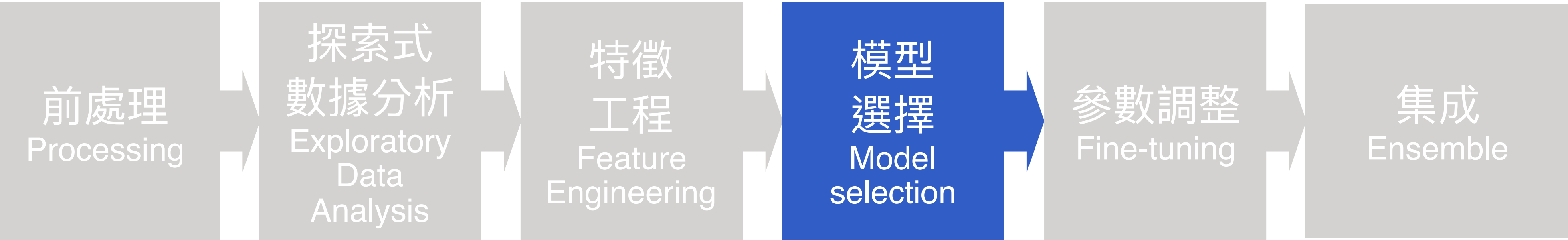
機器學習概論



知識地圖 機器學習 - 模型選擇 機器學習概論

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



模型選擇 Model selection

概論

- 驗證基礎
- 預測類型
- 評估指標

基礎模型 Basic Model

- 線性回歸 Linear Regression
- 邏輯斯回歸 Logistic Regression
- 套索算法 LASSO
- 嶺回歸 Ridge Regression

樹狀模型 Tree based Model

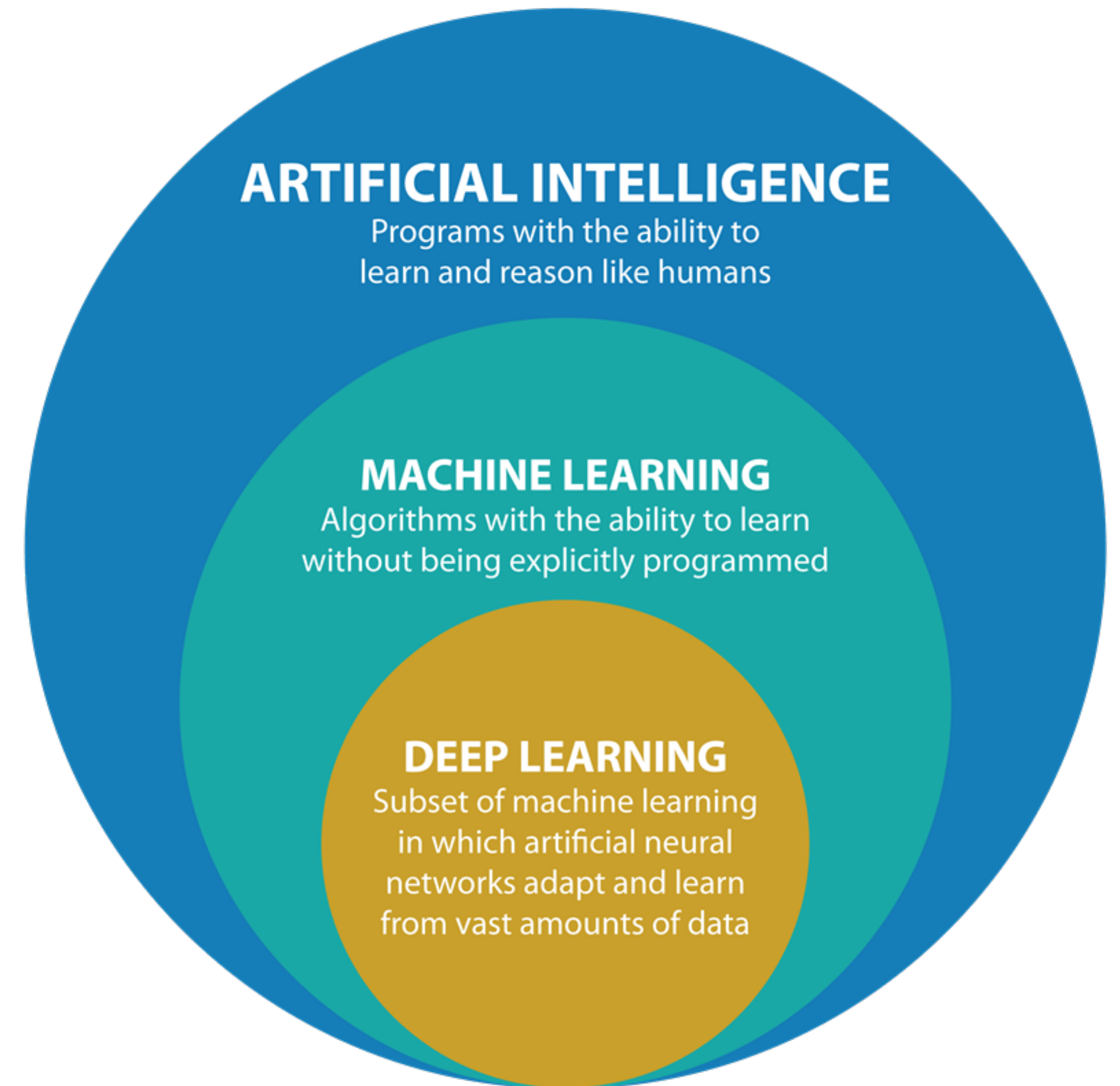
- 決策樹 Decision Tree
- 隨機森林 Random Forest
- 梯度提升機 Gradient Boosting Machine

本日知識點目標

- 了解機器學習與人工智慧的意涵
- 能夠說明機器學習、深度學習與人工智慧之間的差別
- 機器學習中不同領域的意義與應用

機器學習範疇

- 機器學習 (ML) ? 深度學習 (DL) ? 人工智慧 (AI) ? 傻傻分不清楚?
- 機器學習其實是實現人工智慧的技術之一，但因為近幾年機器學習的表現遠超過其他傳統技術，才蔚為風潮
- 深度學習是機器學習中的一個分支，同樣是近幾年的表現遠超過傳統機器學習演算法，才被人們所關注



機器學習是甚麼？

- 白話文：

讓機器從資料中找尋規律與趨勢而不需要給定特殊規則

『 *Field of study that gives computers the ability to learn without being explicitly programmed.*

- Arthur Lee Samuel, 1959



- 數學：

給定目標函數與訓練資料，學習出能讓目標函數最佳的模型參數

機器學習的組成及應用

機器學習有三種! 各自有不同的應用

1. 監督式學習 (常見的應用多屬此類)

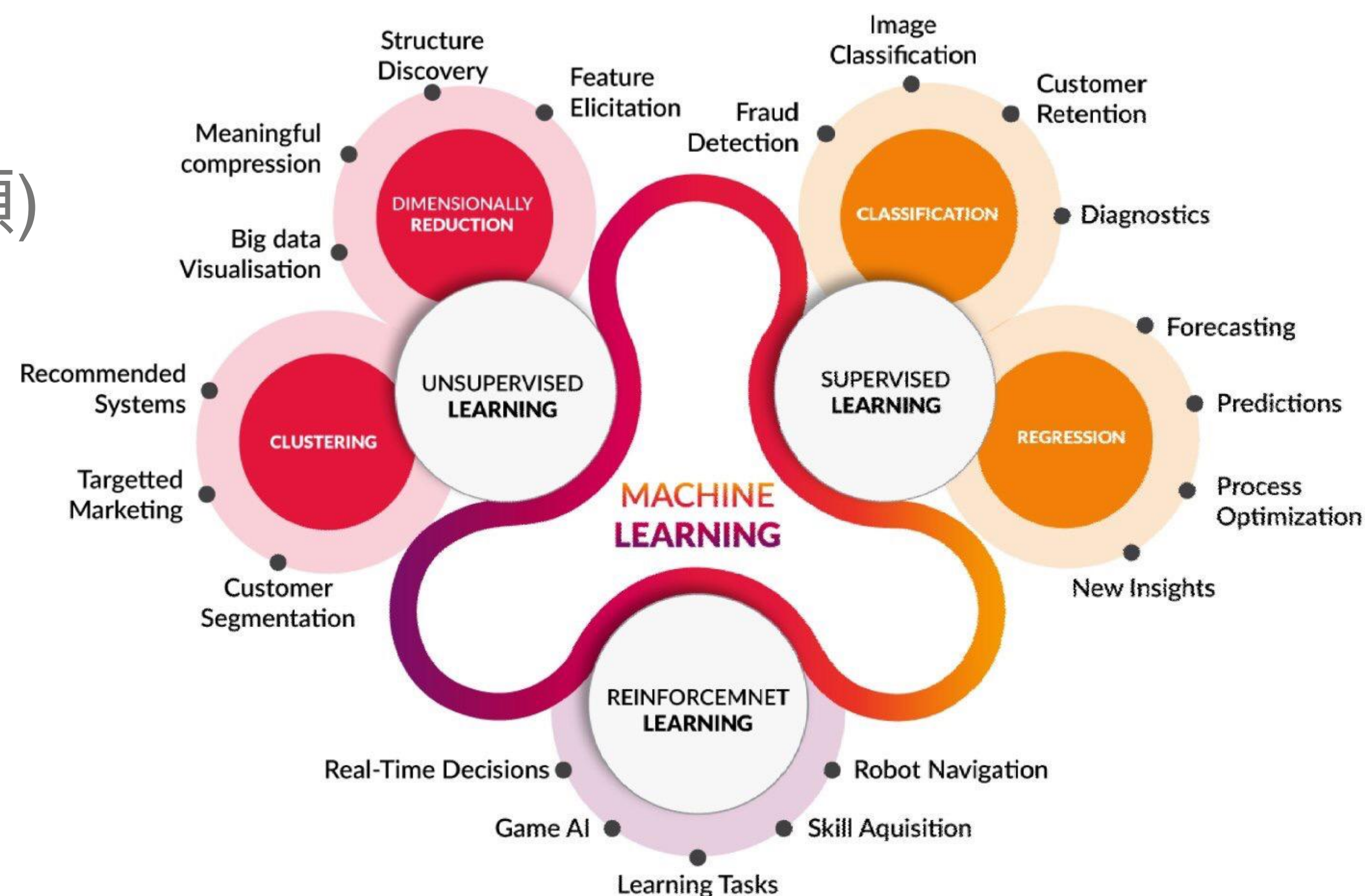
- 圖像分類、詐騙偵測

2. 非監督式學習

- 維度縮減、分群、壓縮等

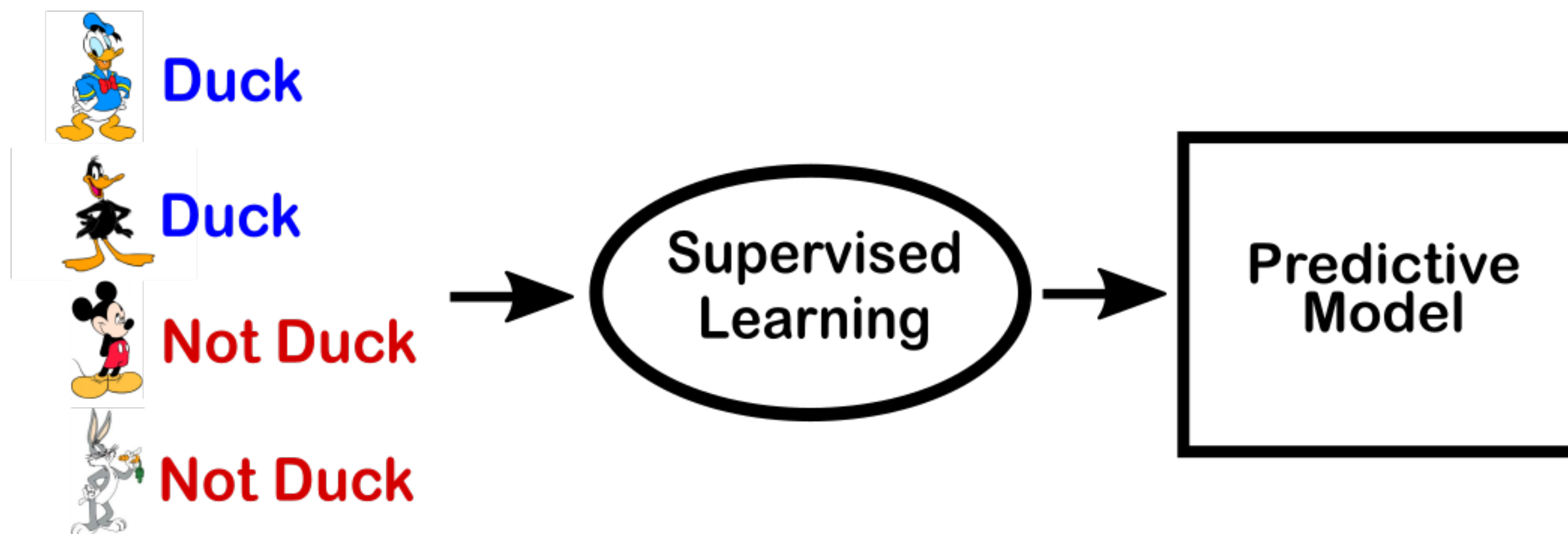
3. 強化學習

- 下圍棋、打電玩



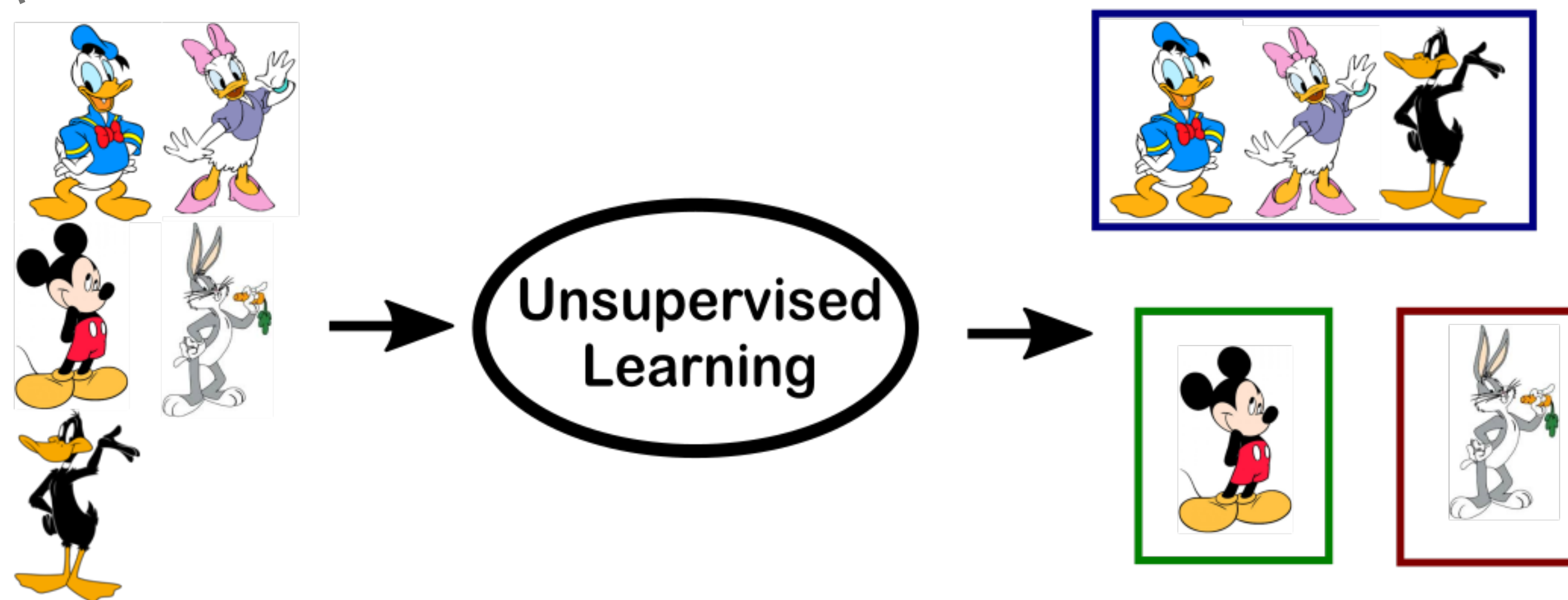
機器學習 - 監督式學習 (Supervised Learning)

- 會有一組成對的 (x, y) 資料，且 x 與 y 之間具有某種關係，如圖像分類，每一張圖都有對應到的標記 (y)，讓模型學習到 x 與 y 之間的對應關係
- 目前主流且有高準確率的機器學習應用多以此類型為主，但缺點是必須要蒐集標註資料



機器學習 - 非監督式學習 (Unsupervised Learning)

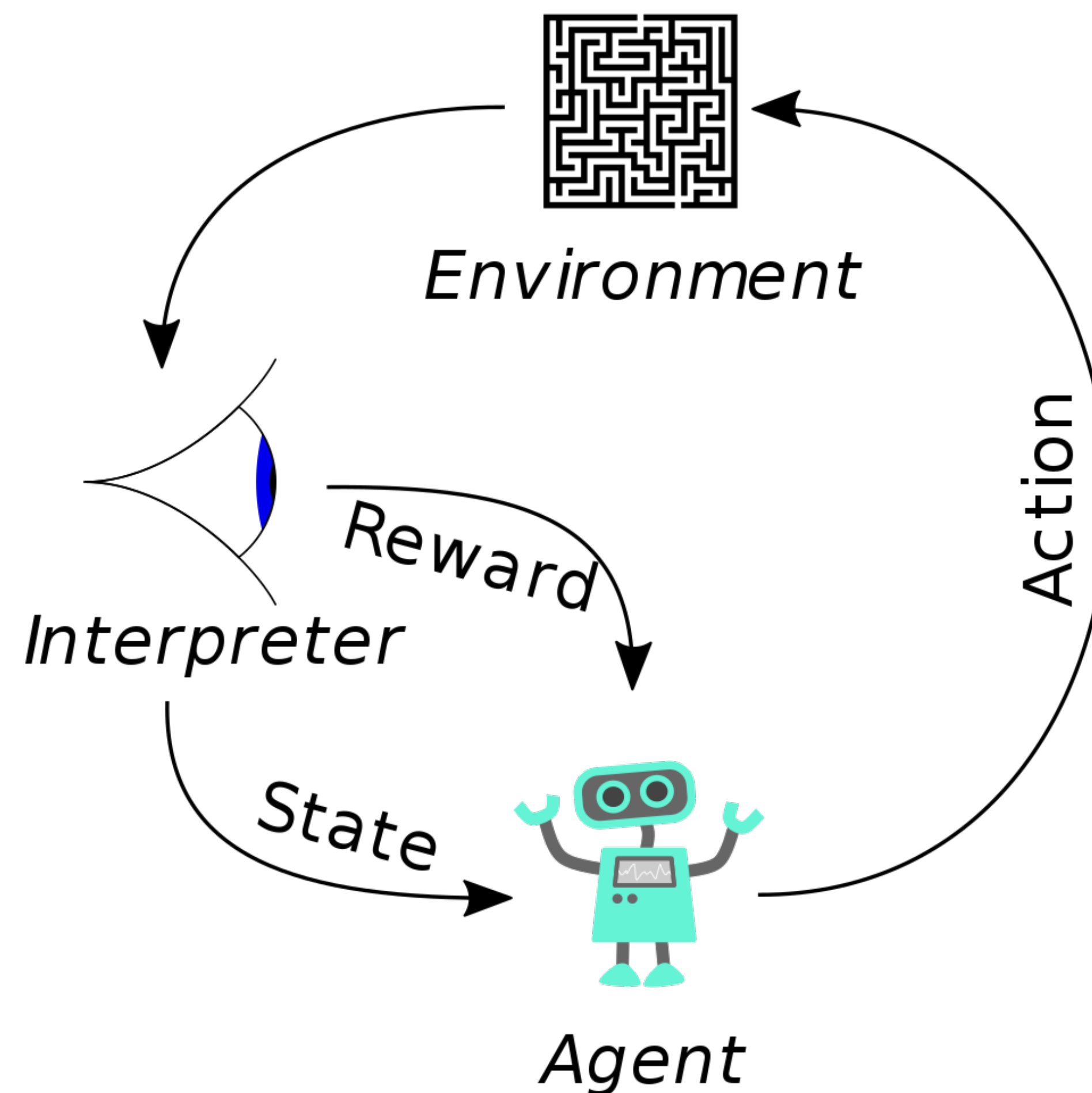
- 僅有 x 資料而沒有標註的 y ，例如僅有圖像資料但沒有標記。
- 非監督式學習通常透過降維 (Dimension Reduction)、分群 (Clustering) 的方式實現
- 非監督式的準確率通常都低於監督式學習，但如果資料收集非常困難時，可應用此方法



圖片來源：[medium](#)

機器學習 - 增強式學習 (Reinforcement Learning)

- 增強式學習是透過定義環境 (Environment)、代理機器人 (Agent) 及獎勵 (Reward)，讓機器人透過與環境的互動學習如何獲取最高的獎勵。
- Alpha GO 就是透過增強式學習的方式訓練，增強式學習近幾年在棋類、遊戲類都取得巨大的進展，是目前非常熱門的研究領域。



常見問題



Q: 機器學習這麼厲害，不就甚麼任務都能夠完成嗎？

A: 的確目前許多題目中機器學習的表現都不遜於人類，但高準確的模型通常需要很大的資料才能達到。另外機器很難學習到語言中的雙關或反諷這些難以數字量化的概念，這也是聊天機器人目前的難題之一。不過針對重複性高、目標明確的問題，機器學習多半都能發揮的不錯。

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

