



# 機器學習簡介



# 機器學習

## 機器學習

- 機器學習\_Wiki

- 人工智慧的一個分支。以「推理」為重點，到以「知識」為重點，再到以「學習」為重點的自然、清晰的脈絡，實現人工智慧的一個途徑之一。

- 電腦可以自動「學習」的演算法。機器學習演算法是一類從資料中自動分析獲得規律，並利用規律對未知資料進行預測的演算法。

- 應用於資料探勘、電腦視覺、自然語言處理、生物特徵辨識、搜尋引擎、醫學診斷、檢測信用卡詐騙、證券市場分析、DNA序列定序、語音和手寫辨識、遊戲和機器人等領域。

- Aws

- Azure

# 機器學習

## 機器學習

- 機器學習是一門人工智慧的科學，該領域的主要研究物件是人工智慧，特別是如何在經驗學習中改善具體演算法的效能。
- 機器學習是對能通過經驗自動改進的電腦演算法的研究。
- 機器學習是用資料或以往的經驗，以此最佳化電腦程式的效能標準。

# 機器學習

機器學習(Machine Learning)簡單來說，就是讓機器去學習，機器要如何去學習呢？

1. 篩選出正確需要的資料
2. 將資料分類
3. 訓練資料(包含訓練及測試兩階段)
4. 將訓練完成的資料模型來預測未來資料

利用python的numpy、pandas可以做許多資料的基本操作，像是可以把蒐集到有缺失值的資料刪除，也可以篩選出某個資料的關鍵屬性...等，接著將資料依據資料的屬性作正確的方式分類，並使用正確的演算法(例如：支持向量機(SVM)、隨機梯度下降(SGD)、決策樹(Decision Trees)...等)再來將分類完成的大量資料分為訓練資料及測試資料。

為什麼要有測試資料呢？因為在資料訓練上，訓練模型只會越來越好(訓練的精確值越來越高)，為了避免過度擬合(Over fitting)的狀況，可以利用測試資料來確認，訓練完成的資料模型的測試出來的資料是否值得信任，最後，就可以利用訓練完成資料模型來預測未來的數據。

# 機器學習

機器學習中，教電腦進行預測或推理。

- 首先，可以使用演算法和範例資料來訓練模型。
- 然後，可以將模型整合到應用程式中，以便即時且大規模地產生推論。

**監督式學習**演算法訓練範例資料，以指定演算法的輸入和輸出。

- 例如，資料可能是手寫數字的影像，這些影像會加上註解，以指示其所代表的數字。
- 只要有足夠的標記資料，監督式學習系統最終會辨識出與每個手寫數字關聯的像素和形狀叢集。

**非監督式學習**演算法使用未標記的資料來進行訓練。

- 這些算法會掃描新資料，並在未知的輸入值與預先確定的輸出值之間建立有意義的聯繫。例如，
  - 非監督式學習演算法能夠將來自不同新聞網站的新聞報導按照相同的類別進行分組，
  - 例如運動類新聞和犯罪類新聞。

# 機器學習

## 監督學習

- 從給定的訓練資料集中學習出一個函數，當新的資料到來時，可以根據這個函數預測結果。監督學習的訓練集要求是包括輸入和輸出，也可以說是特徵和目標。訓練集中的目標是由人標註的。常見的監督學習演算法包括迴歸分析和統計分類。
- 監督學習和非監督學習的差別就是訓練集目標是否有人為標註。他們都有訓練集 且都有輸入和輸出。

無監督學習與監督學習相比，訓練集沒有人為標註的結果。常見的無監督學習演算法有生成對抗網路（**GAN**）、聚類。