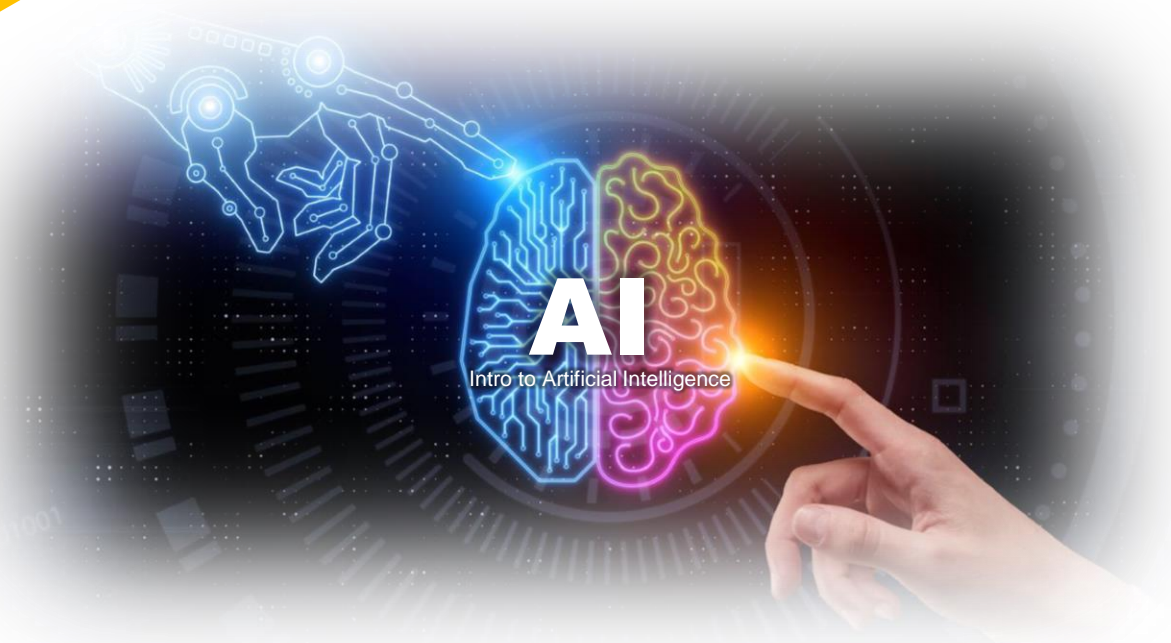


教學/講義製作：
社長 王修佑

人工智慧 導論



政附資訊16th 社課VIII

2022/03/11

王修佑 hugo.tw.wang@gmail.com

Table of Contents

01

人工智慧簡介

Intro to Artificial Intelligence

02

監督式學習

Supervised Learning

03

模型評估

Model Evaluation

04

深度學習

Deep Learning

05

非監督式學習

Unsupervised Learning

06

人工智慧與遊戲

Applications of Artificial Intelligence

人工智慧 簡介

01

- ◆ What is Artificial Intelligence?
- ◆ 人工智慧的應用
- ◆ 特徵 Feature

What is Artificial Intelligence?

發展目標

讓機器能呈現甚至超越人類智慧的技術。

A machine that can sense, reason and act.

使用的優點



不會疲倦



成本低



Processing Big Data

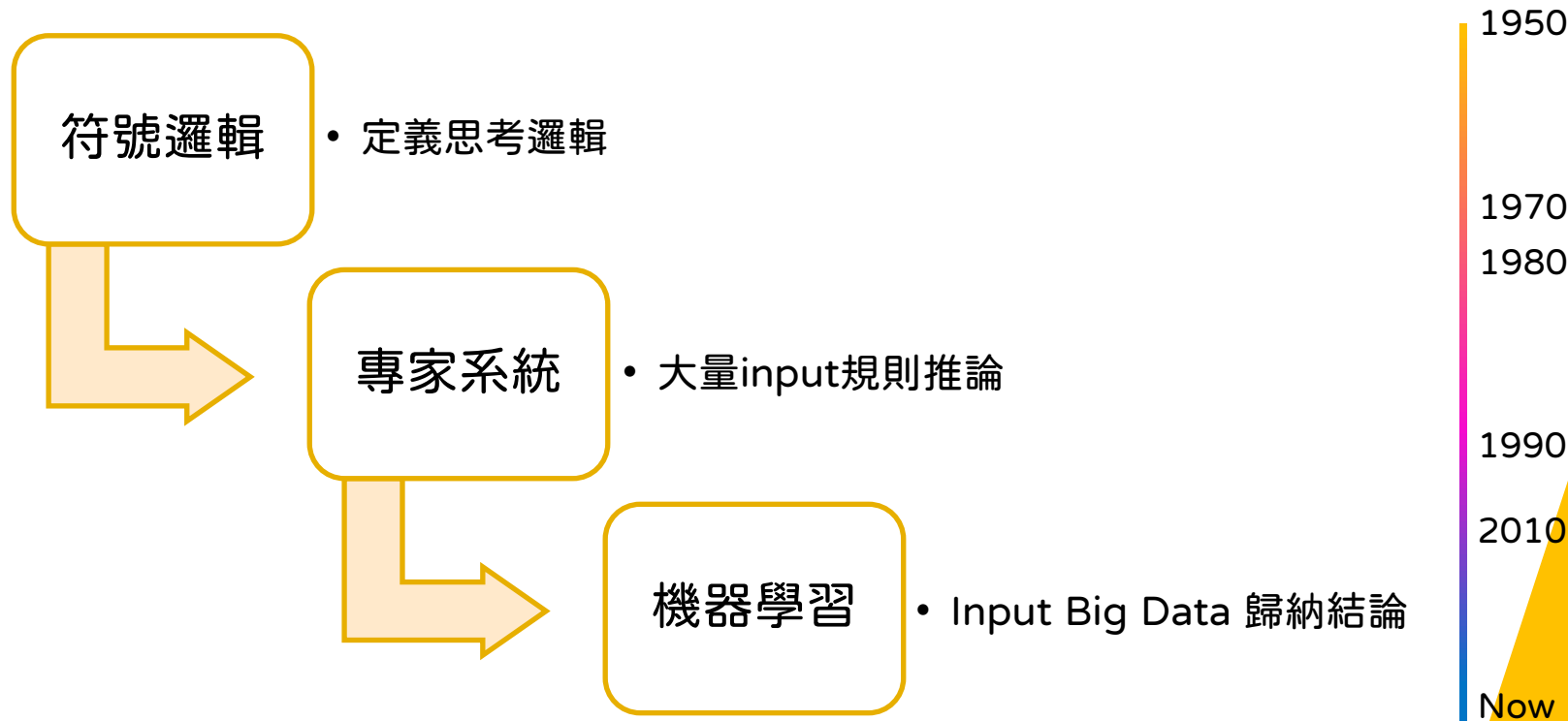


速度快



可在多元環境工作

簡史

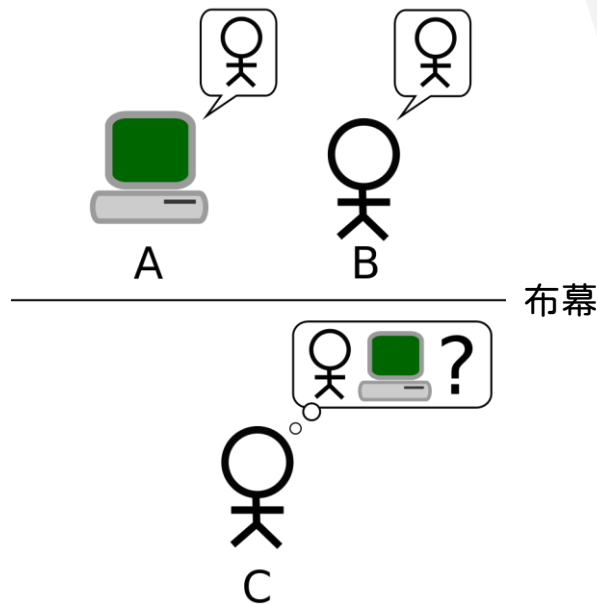


起源

圖靈測試實驗

定義與顯示是否具備”智慧”

1950年 - Alan Turing提出圖靈測試理論



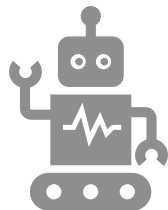
人工智慧的種類

👉 Strong AI :

與人類相同的完整知識能力，像是推理。

👉 Weak AI :

具備特定領域範圍能力。



Machine Learning

💻 傳統程式：寫出如何達到目標的邏輯。

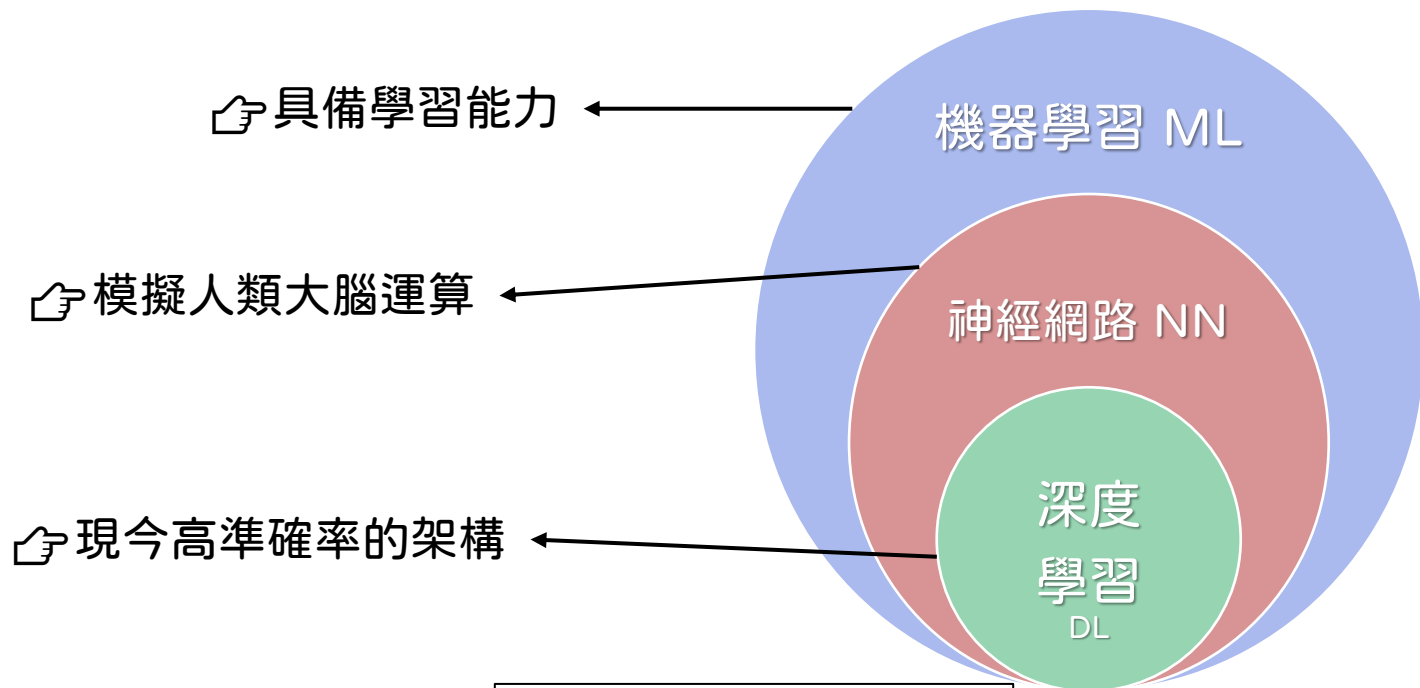
🤖 Machine Learning：像人類般的學習能力。

給學習的方式，
累積經驗

“A field of study that gives machines the ability to learn without explicitly programmed.”

- Arthur Samuel (1959)

Deep Learning



機器學習基本方法

監督式學習

Supervised Learning

- 給機器data的正確答案，以逐漸拉近預測的正確性
- 學習目標：函數 f

非監督式學習

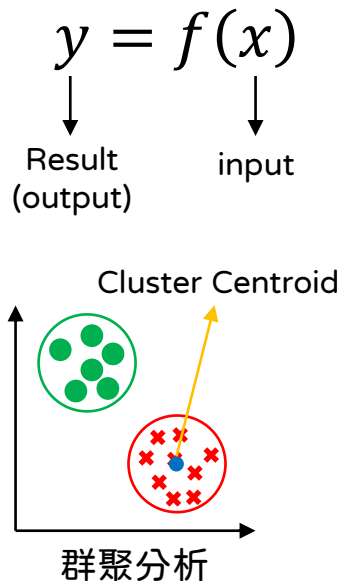
Unsupervised Learning

- 機器不知道data的意義
- 相似程度比較 ㄅ 物以類聚

強化學習

Reinforcement Learning

- 以行為(action)的好與壞，不斷嘗試更好的結果
- Skinner Box



人工智慧的應用



🕒 解決問題

技術與應用

迴歸 (regression)

Supervised Learning

找出data的規律

result

預測

透過函數“ f ”

👉 結果方法“迴歸”到問題；且
 $f(x) = y$ 的“ y ”具物理意義。

技術與應用

分類(classification)

Supervised Learning

分類輸入已定義好類別的樣本

歌曲
辨認

input



f

辨識



Lady Gaga 1



Jason Marz 2



Ed Sheeran 3

output

👉 輸出 $f(x) = y$ 的“ y ”為代表編號；且“ y ”不具物理意義。

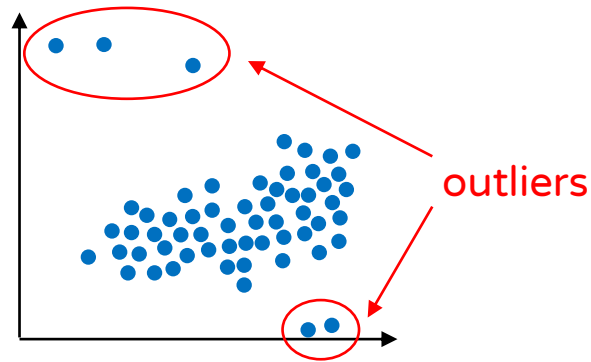
技術與應用

分群(clustering)

Unsupervised Learning

- 尋找關聯性特徵
- 可挖掘data中”潛在特性”
- 例如：推薦系統、其他人也喜歡功能、網購平台、(推薦)你可能喜歡、異常分析

⚠異常分析(駭客攻擊)



技術與應用

電腦遊戲(gaming)

- 模擬非玩家角色 (NPC)
 - 👉 自動尋路
 - 👉 決策分析
 - 演算法設計
- ↓
- Machine Learning
 - Google DeepMind ALPHAGO
 - DeepMind Alphastar
 - 強化學習與圍棋



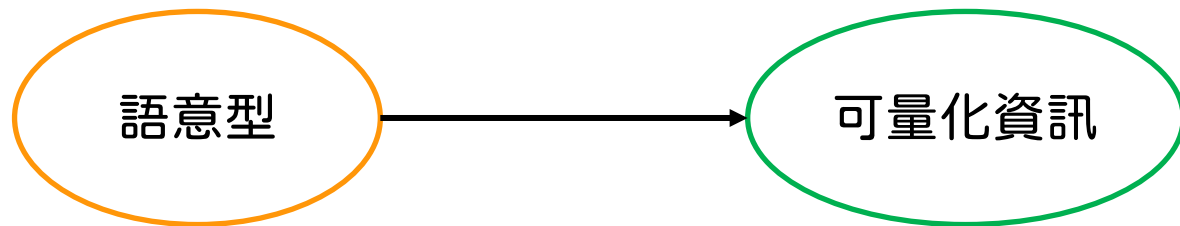
00 特徵



09 什麼是特徵?

- 區分不同物理的依據
- 統整特徵做出分類 >>> 「分類器」
- ML的目標：找出特徵，學習分類器行為。
- ㄟ 精準決策

機器看特徵



Baseball vs. Basketball

Train vs. plane

👉 機器不易理解

球的直徑

球的極速

📖 同樣的特徵解決不同問題 >> 不同的結果

特徵

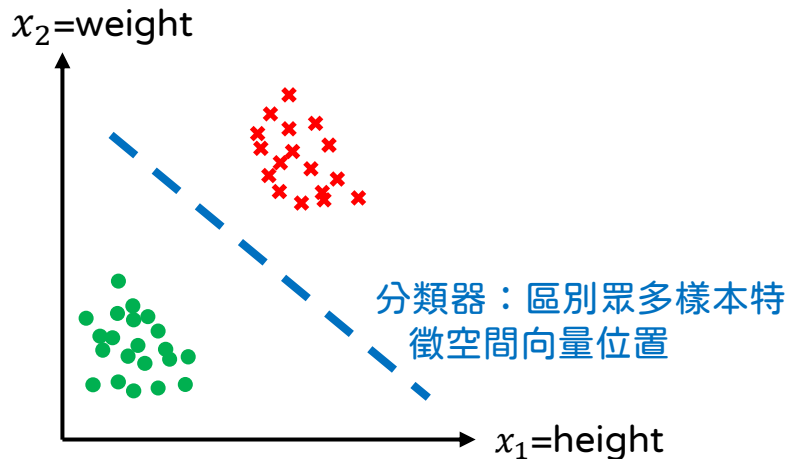
✿ 使用多個特徵

- 因較複雜的問題而產生
- 例如：身高178 cm; 體重78 kg

↓
向量表示法(178, 78)

↗
特徵向量

✿ 特徵空間

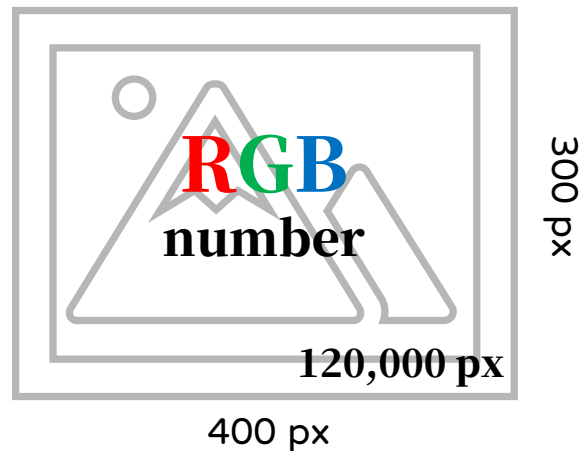


電腦如何理解影像

👁️ 人看到的



💻 電腦看到的



影像特徵

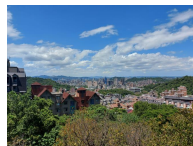
<彩色>



<黑白>



影像特徵



<彩色>



<黑白>

- 1 pixel >> (R,G,B)

- 亮度

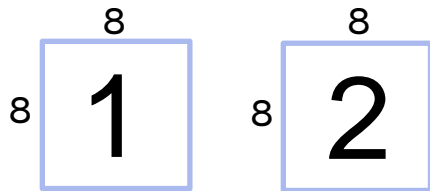
☞ 數字辨識：0、1、2……

8x8 image



數位化/平面化

1x64 image



210 | 87 | | 220 | 225

監督式 學習

- ◆ 資料整理
- ◆ 迴歸分析
- ◆ 梯度下降法與學習率
- ◆ 淺談分類-以Titanic dataset應用為例

02

資料整理的重要 – 1.蒐集資料

- 需大量Data

傳統
數千、百筆



Now
數萬筆以上

- 物件辨認

例如：IMAGENET上提供1,000,000張images資料，辨識正確率97%以上。

- 來源

自行蒐集：問卷、IoT設備、網路爬蟲

公開資料：<國內>政府資料開放平台 <國際>kaggle、UCI ML library

2. 資料整理



數據補遺

方法1：內插與外插法

方法2：取平均數、中位數

方法3：補零（使機器忽略不記）

2. 資料整理

數據正規(量化)

(明顯非數字) 性別：male/female

(物理意義不同) 風力：0~14級

(物理意義不同) 氣溫：-50~50°C

數據化



固定的範圍 (有限)

0, 1

$0, \frac{1}{14}, \frac{2}{14}, \frac{3}{14}, \dots, \frac{13}{14}, 1$

$0, \frac{1}{100}, \frac{2}{100}, \frac{3}{100}, \dots, \frac{99}{100}, 1$

- 多種可能性或化為數字較不具意義的資料 (例：科系、形狀……)

👉 使用one-hot encoding量化 >> 單一位置=1，物理意義變成「是否為某特定shape」

例如： circle = [1, 0, 0, 0]

triangle = [0, 0, 1, 0]

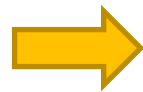
square = [0, 1, 0, 0]

rectangle = [0, 0, 0, 1]

2. 資料整理

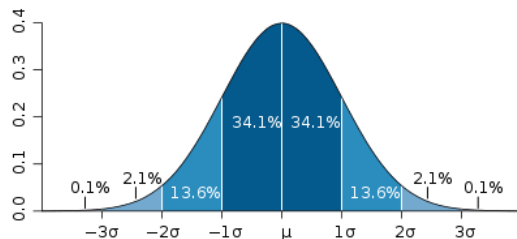
數據異常

- 特殊原因產出「不合常理的值」



刪除

去頭去尾法：固定delete資料分布的前後固定百分比資料。



- Data一般呈現高斯分布

☞ 去除超過2~2.5個標準差外的資料

= 保留95%的Data給機器學習

數據洩漏

- 訓練模型使用了不該使用的特徵

☞ 可能與與預測結果高度相關，但實際應用沒有絕對意義

☞ 發現正確率過高：模型特徵與預測結果高度相關(或有時是已知的影響因子)