

# 關於機器學習

107學年度第1學期

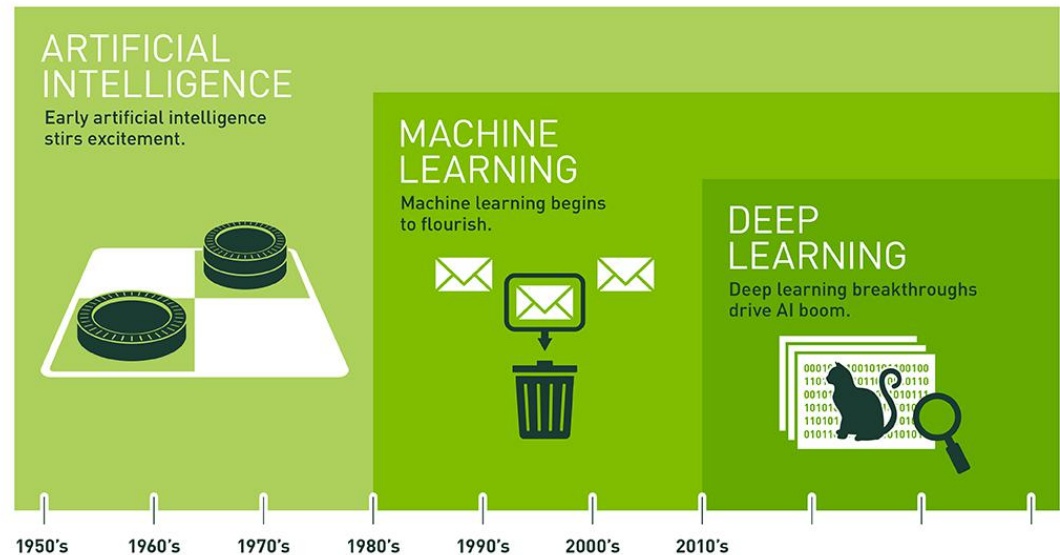
財務金融系 張子溥 助理教授

E-mail: [changtp@yuntech.edu.tw](mailto:changtp@yuntech.edu.tw)



# 教學目標

- 本課程將教授常見的機器學習方法，並且以金融實例及資料作為範例，讓學生實作機器學習技巧。透過本課程讓學生了解機器學習如何應用在金融領域，也能夠對於金融科技發展趨勢有更進一步的了解。



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# 教材

序號	教材類別	編訂方式	書名/ISBN	作者	出版者	出版年
1	教科書	非自編	資料挖礦與大數據分析	簡禎富、許嘉裕	前程文化	2014
2	參考書	非自編	R資料採礦與數據分析	何宗武	碁峯	2016
3	參考書	非自編	R語言：數學計算、統計模型與金融大數據分析	鄧士昌	博碩	2016





# 橫掃網絡的神秘圍棋高手Master

## Master 主要手下敗將



用梓 立心勸九叔，勸哈咁係「all forces」

**日本** 井山裕太九段、六冠，網路暱稱「jngo01」

資料來源：台灣《蘋果日報》



# 誰會贏美國大選？過去命中三次的 AI，這次押川普！

2016/10/30

- 一個名為 MogIA 的 AI 系統從 Google、Facebook、Twitter 和 YouTube 上收集超過 2000 萬個數據點並進行分析後，預測川普將會獲得今年美國大選最終勝利。MogIA 由印度新創公司 Genic.ai 開發，此前已經成功預測二屆總統大選結果。
- 自 2004 年研發完成後，MogIA 憑借機器學習技術不斷提高數據分析能力，變得越來越「聰明」。參與到關注度最高的總統大選預測中，但 MogIA 算是其中經驗最豐富、準確率最高的「老手」。
- 與 MogIA 的預測相反，目前希拉蕊在民調支持率中領先，但據 MogIA 預測的支持率變化趨勢，川普會在最後時刻領先，最後勝率甚至已經比 2008 年時的歐巴馬高出 25%。





# 2018 十大科技趨勢預測

## 第一大類：智慧

### 趨勢1 人工智慧與機器學習

- 由多種科技與技術組成。繼**2016**年超級電腦**AlphaGo**打敗韓國棋王後，激起全球人工智慧（**AI**）狂潮和對機器學習（**Machine Learning**，**ML**）的討論。未來將進入能理解、學習、預測，甚至自行運作的系統。西爾利指出，**AI**和**ML**造就各種智慧應用，硬體如機器人、自駕車和消費性電子，在軟體服務方面，則有虛擬個人助理（**Virtual Personal Assistant**，**VPA**）和智慧顧問等。

### 趨勢2 智慧應用程式

- 未來，虛擬個人助理將是最常出現的應用程式，具備真人助理的部份功能，讓日常工作更有效率，譬如識別出電子郵件中的重要內容。另一種就是虛擬顧客助理，在特定領域內提供服務，譬如銷售與客服。這類智慧應用程式（**Intelligent App**）未來將可能改變工作本質與職場結構。

### 趨勢3 智慧物件

- 物件能按固定的程式模型執行任務，還能利用人工智慧和機器學習做更進階的行為，同時用更自然的方式和周遭環境及人類進行互動。未來，當無人機、自駕車和智慧家電等智慧物件（**Intelligent Things**）普及後，我們預測，智慧物件的運作模式將從現在的獨立運作，進階成協作模式。



- 第二大類：數位

- 趨勢4 虛擬實境與擴增實境
- 趨勢5 數位分身
- 趨勢6 區塊鏈

- 第三大類：網絡

- 趨勢7 對話式系統
- 趨勢8 網狀應用程式和服務架構
- 趨勢9 數位科技平台
- 趨勢10 適應性安全架構





# 什麼是人工智慧 ( Artificial Intelligence ) ?

- 希望電腦能夠擁有智慧 ( Strong AI)
- 希望電腦能夠展現出有智慧的外顯行為 (weak AI)

The image shows a Pepper robot standing next to a man in a suit. The robot is white and has a tablet on its chest displaying a blue bow tie. A large red speech bubble contains the following text:

**哈囉！我叫Pepper**

- 我來自日本
- 身高121公分，體重29公斤
- 月薪26888元，9月下旬來台
- 雇我要綁約24個月

我精通中文、日文、英文〈可設定其他語系〉

我專長是迎賓接待、產品推薦、行銷互動、資訊蒐集〈總之我很好用〉

體力狀況  
(電池續航力) / 8-12小時

靈活度 / 擁有20個可動關節

情緒偵測 / 27個感應器

Pepper 來台求職，鴻海副總裁呂芳銘 (左) 宣布面試成功。  
記者胡經周／攝影



# 人工智慧已經有一段歷史

- 第一次熱潮：1950-1960年代，運用電腦針對特定問題進行推論與探索。要證明難解的定理或是西洋棋對弈，對電腦而言相當地簡單；但現實生活的問題，卻無法解決
- 第二次熱潮：1980年代，把「知識」輸入到電腦裡，電腦就會變聰明→專家系統
  - 如何翻譯 He saw a woman in the garden with a telescope.
  - 他 看到 一位女性 在花園裡 用望遠鏡
  - 知識多到無法完全表達出來

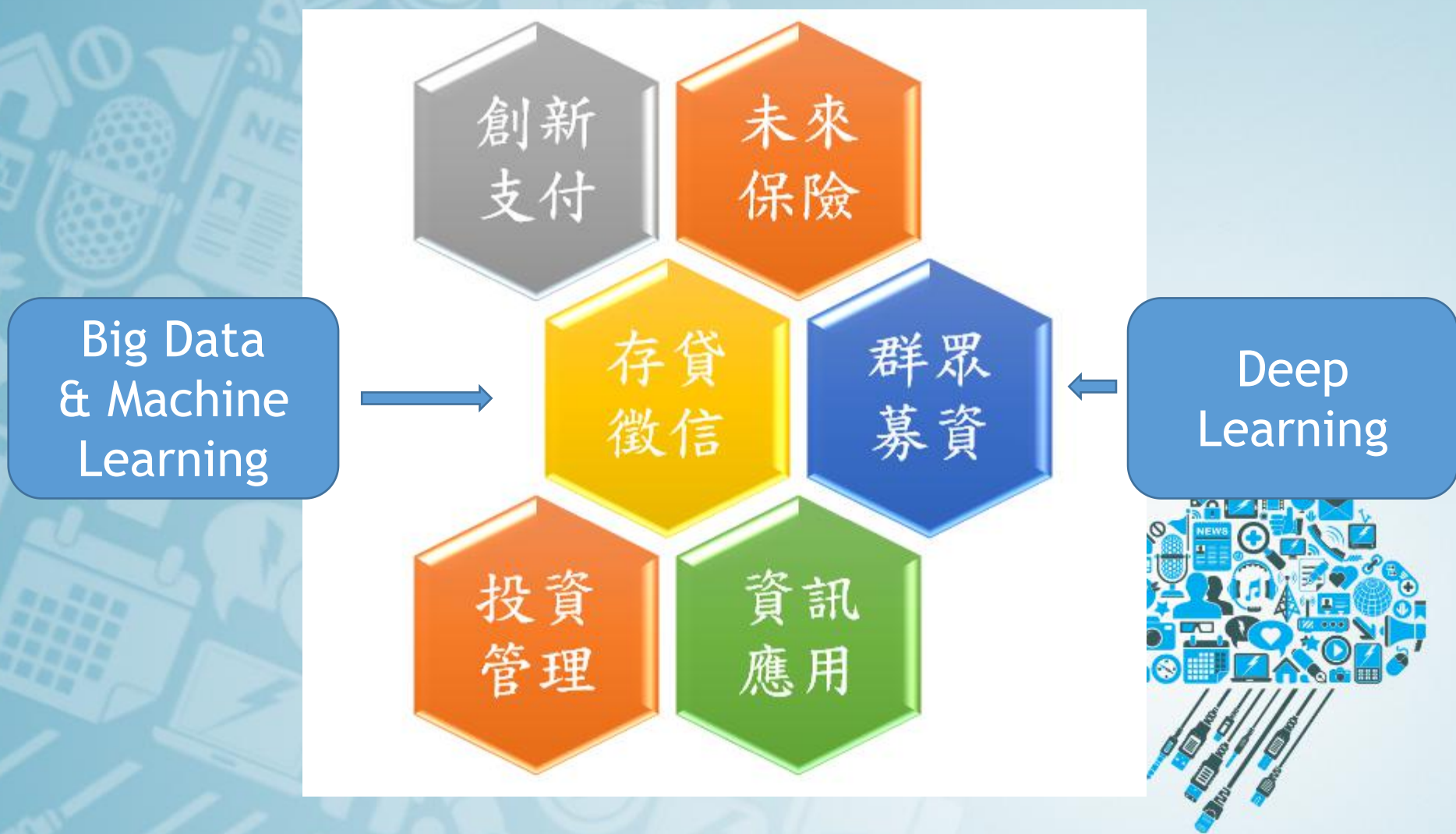


# 我們正在AI的第三次熱潮上

- 多數學者放棄強人工智慧，往弱人工智慧靠攏
  - 搜尋引擎、語音辨識
- **INTERNET**的崛起，資料蒐集變得容易
- 電腦計算能力非比從前，很多過去無法實現的演算法變成可行
  - 機器學習、深度學習



# 人工智慧(AI) 與 金融科技(FinTech)





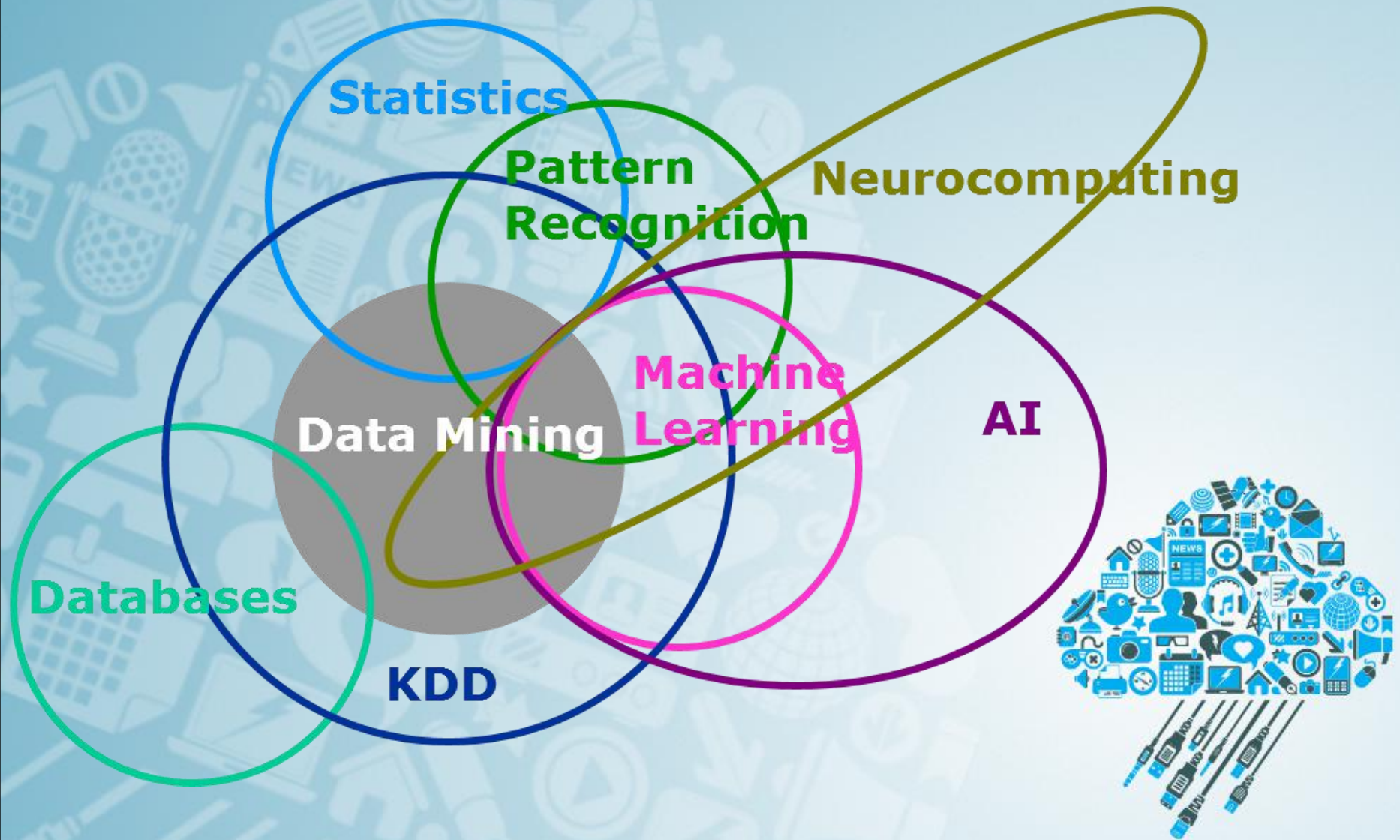
# 資料探勘、機器學習和人工智慧，您認為最大的差異是什麼？能否用例子來說明上述三者的差異？

- 人工智慧是一種系統透過資料處理和(或)演算法運算傳達情報資訊的方法，從而產出有意義的資訊，並讓機器擁有比人更聰明的智慧。簡單來說，人工智慧就是教電腦做「正確、聰明的決策」，也就是我們所謂的「智慧」。
- 而機器學習是人工智慧的其中一個分支，讓機器可以自動學習、從巨量資料中找到規則，進而有能力做出預測。人工智慧讓過去只能透過人類或動物智慧解決的問題也能透過電腦系統迎刃而解；機器人是自動執行工作的機器裝置，而人工智慧則可以讓機器人快速、精準處理大量資料。簡單來說，機器人像是人的「身軀」，人工智慧則是人的「腦」。
- 資料探勘是利用分析技術來發掘資料間未知的關聯性與規則，而人工智慧是其中的一個分析技術。傳統會用統計的方式來做資料探勘，現在也有些人會用機器學習的方法來處理大量資料。由於機器學習可處理巨量多維度資訊，並能發掘多元變動因素之間的關聯性，再藉由優化實現清晰與可辨識的目標，因此非常適用於資料探勘。

Appier沛星互動科技  
創辦人兼執行長游直翰



# 資料探勘、人工智慧、機器學習





# 你要如何教電腦這是一隻貓







# 先學習「分類」再用來「預測」

- 所謂的機器學習，是一種人工智慧的程式自行學習的機制
- 而構成學習的基礎元素是「分類」這樣的處理動作。只要能做好分類，既能理解事物，也能做出判斷、採取行動。
- 機器學習可以讓電腦一面處理大量的資料，一面自動學會這樣的分類方式。等到學會分類方式後就可以用來預測未知的資料。





# 機器學習案例 —「分類」

我是特徵					我是目標
菇種	菌傘形狀	菌傘顏色	分布地帶	氣味	有毒 / 無毒
菇菇 A	球狀	棕色	腐木	腥味	0 (標籤)
菇菇 B	圓錐形	淺黃色	草堆	無氣味	1 (標籤)
菇菇 C	圓錐形	白色	樹葉	霉味	1 (標籤)
菇菇 D	鐘形	紫色	腐木	杏仁味	0 (標籤)
菇菇 E	下凹形	黃色	腐木	無氣味	1 (標籤)
菇菇 F	扁平狀	白色	泥土	惡臭味	0 (標籤)

當一筆新資料輸入電腦中，比如特徵具備白色鐘形菌傘、分布在腐木上、杏仁味的香菇，電腦即會判斷這朵香菇有毒或沒毒的機率有多高了。





## Learning

**Existing Data**



**Machine (Algorithm)**



**Model**

## Prediction

**New Data**



**Model**



**Prediction**



# 換一種說法

觀察



學習



獲得技能

資料



機器學習

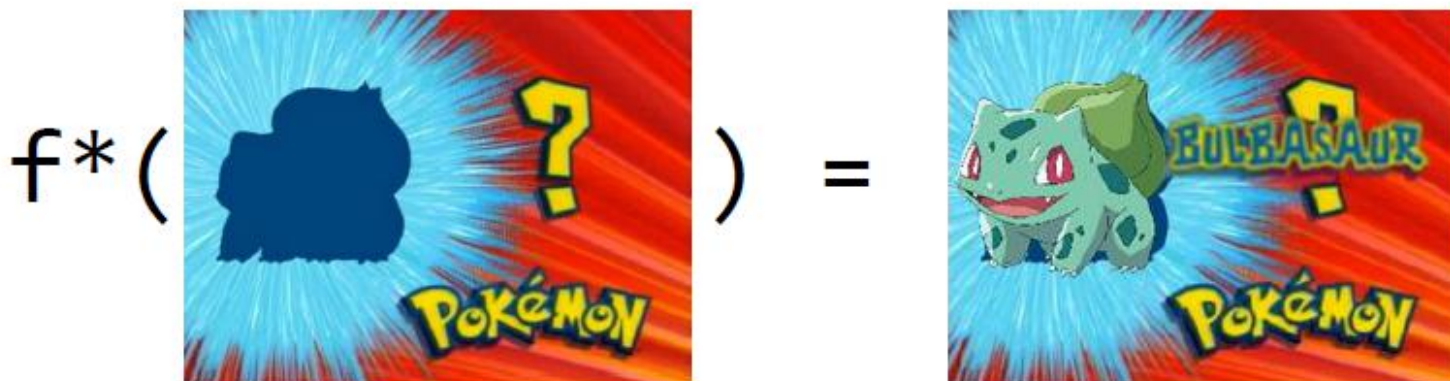


獲得技能

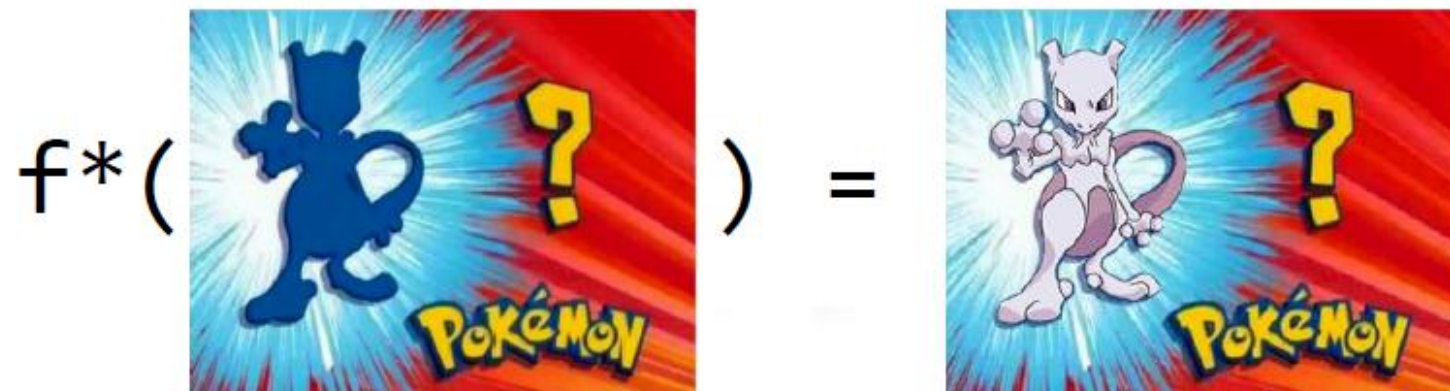
預測得更  
準，績效越  
好



# 每集寶可夢結束的“猜猜我是誰”



Ans: 妙蛙種子



Ans: 超夢





# 沒有學習到的結果

$f_1($



) =



$f_1($

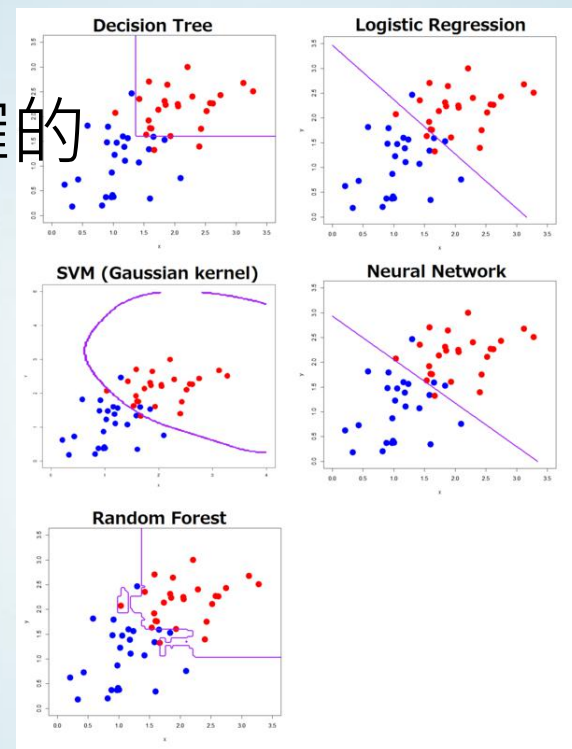


) =



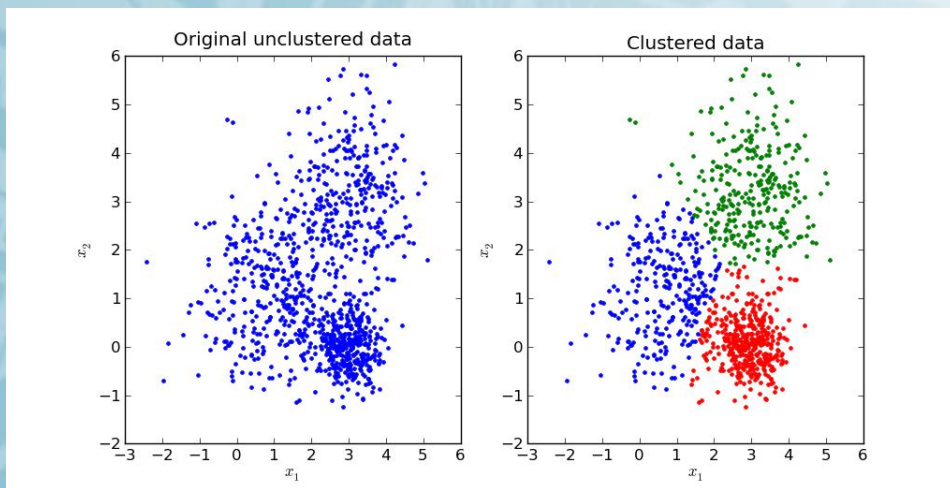
# 兩種機器學習的方式 (I)

- 監督式(supervised)學習，也就是「有老師的學習」
- 給電腦準備好的資料(輸入)與正確答案(輸出)
- 人類扮演老師的角色，教導電腦正確的分類方式。
- 分類與迴歸



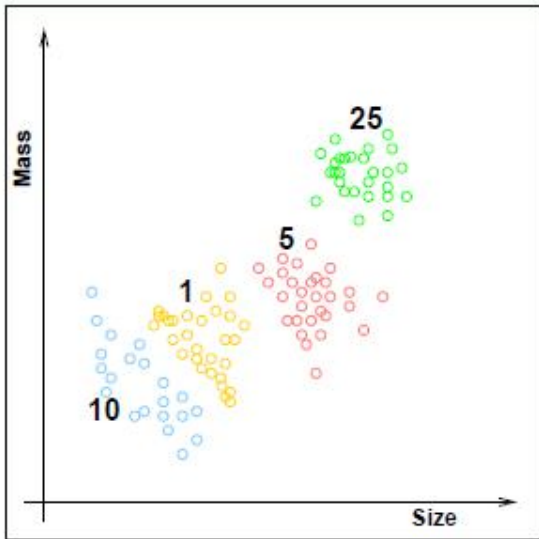
# 兩種機器學習的方式 (II)

- **非監督式(unsupervised)學習**，也就是「沒有老師的學習」
- 只給電腦輸入用的資料，讓電腦找出隱藏在資料中的結構或規則
- 分群與關連規則

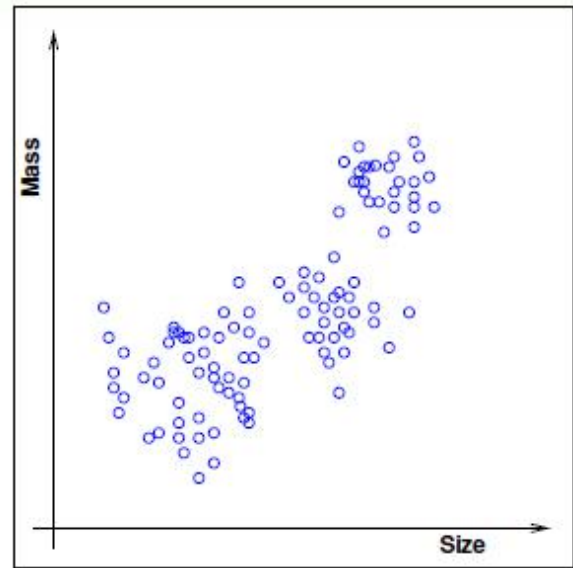




# 辨別硬幣



supervised multiclass classification



## unsupervised multiclass classification



# 監督式學習

- K-最近鄰法
- 決策樹
- 隨機森林
- 支持向量機
- 簡單貝氏分類
- 類神經網路
- 迴歸(羅吉斯迴歸)
- LDA、QDA



# 非監督式學習

- 群集分析 (Clustering)
- 關聯規則學習
- 資料維度縮減 (PCA、Factor analysis)





# 上課進行方式

- 輕鬆的
  - 介紹ML與AI在各行各業的應用
- 頭昏的
  - 聽我講一堆原理，盡量減少數學與統計
- 動手的
  - 用R來跑跑常見的機器學習方法

