



SCU LawTech

Class #08

人工智慧 AI ARTIFICIAL INTELLIGENCE

機器展現出人類般的智慧

機器學習 ML MACHINE LEARNING

使機器達到人工智慧的
其中一種方法

深度學習 DL DEEP LEARNING

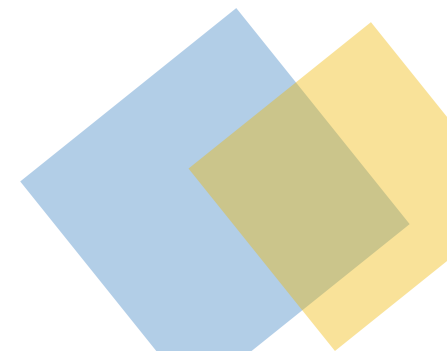
實現機器學習的其中一項技術

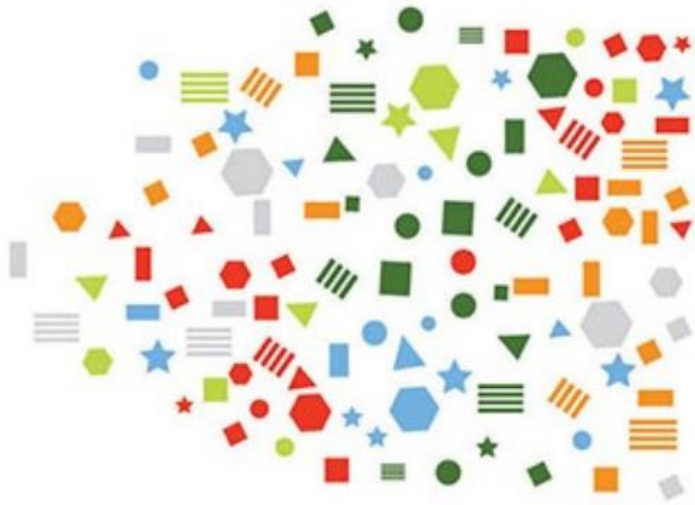
● 1950年 ● 1960年 ● 1970年 ● 1980年 ● 1990年 ● 2000年 ● 2010年 ● ● ● ●

<https://medium.com/@hiskio/%E7%94%9F%E6%88%90%E5%B0%8D%E6%8A%97%E7%B6%B2%E8%B7%AF%E5%88%B0%E5%BA%95%E5%9C%A8gan%E9%BA%BB-f149efb9eb6b>

數據分析的過程 – 傳統 v.s. 大數據

- 掌握資料
- 問題的理解
- 分析的思路、流程
- 結果的解讀
- 決策





Unorganized Transactional Data

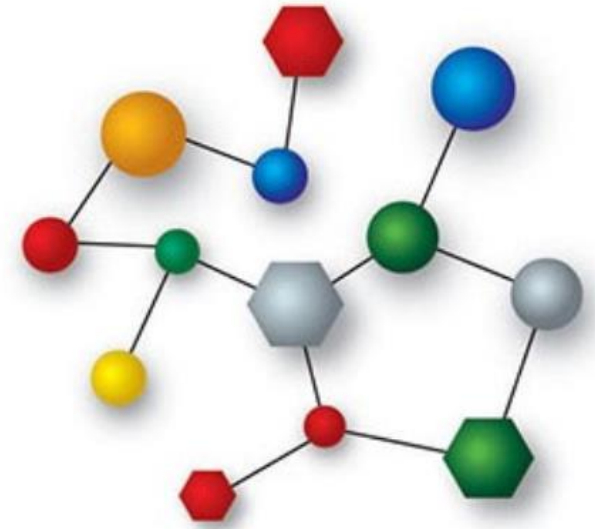
1



```
1 0 1 0 1 0 1 0 1 0 1 0 1
1 1 1 1 1 0 1 0 1 1 1 1 1
1 1 0 0 1 0 1 0 1 1 0 0 1
1 0 1 0 1 0 1 0 0 0 1 1 0
1 0 1 0 1 0 1 0 1 0 1 0 1
0 1 0 1 1 0 1 0 1 1 0 0 1
1 0 1 1 1 0 1 0 1 1 0 1 0
1 1 0 0 1 0 1 0 1 0 1 0 1
```

Data Processed by the Algorithm

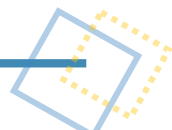
2



Intelligent Associations

3

資料類型



資料型態有哪些？

資料衡量尺度	資料型態	特性
文字資料	非結構化型態	無法直接進行運算
名目資料	質化型態	類別
順序資料	質化型態	優先順序
區間資料	量化型態	大小距離
比例資料	量化型態	可進行四則運算的數值

資料可能發生的情況



- 缺失值
- 異常值
- 不一致的值
- 重複數據及含有特殊符號（#、*、\$）的數據

缺失值的原因及影響



原因

- 有些訊息暫時無法獲取，或是獲取信息的代價太大
- 被遺漏的，可能忘記填寫、採集設備故障、傳送時遺失...等
- 屬性值不存在，不一定表示數據有錯誤

影響

- 建模將丟失大量的有用信息
- 建模所表現出的不確定更加顯著
- 使建模過程陷入混亂，導致不可靠的輸出

Pandas 兩個主要的資料結構

- Series：單維度的資料
- DataFrame：多維度的資料

	年度	字號	案號	類型	裁判日期	裁判案由
0	109	台上	1719	刑事判決	109.04.15	違反證券交易法等罪
1	108	台上	4056	刑事判決	109.04.15	違反證券交易法等罪
2	108	台上	16	刑事判決	109.01.16	違反證券交易法等罪
3	109	台抗	46	刑事裁定	109.01.09	違反證券交易法等罪不服再執行羈押
4	107	台上	846	民事判決	108.03.27	請求損害賠償

