

資料模式與資料建模

大數據分析與AI概念

國立中興大學資訊工程學系
陳奕中副教授
chenyich@nchu.edu.tw





- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：設計新的大數據題目



何謂大數據分析？

常見的定義1：大數據分析是分析大型複雜數據源以發現趨勢、模式、客戶行為和市場偏好，從而為更好的業務決策提供信息的過程。...

常見的定義2：大數據有4種，稱為4V：Volume大，資料量大數據與傳統數據最大的差別在於資料量，資料量遠大於傳統數據，因此以「大數據」一詞來敘述並區分兩者...

簡單的說：

1. 定義你要解決的問題
2. 收集與這問題有關的所有資料
3. 把所有資料交給電腦，由電腦來找出目標問題的規則
4. 直接在應用中使用這些規則或加值這些規則(迴歸、預測)

- 大數據分析領域流傳最廣泛的介紹影片
 - Kenneth Cukier: Big data is better data
(<https://www.youtube.com/watch?v=8pHzROP1D-w>)



何謂大數據分析？

➤大數據分析功用1：協助我們找出問題隱藏規則



乳癌辨識原本僅知9種特徵，但透過大數據分析找到3種新的特徵，且這3種新特徵辨識效果較原本9種好。

*照片來源: 康健雜誌



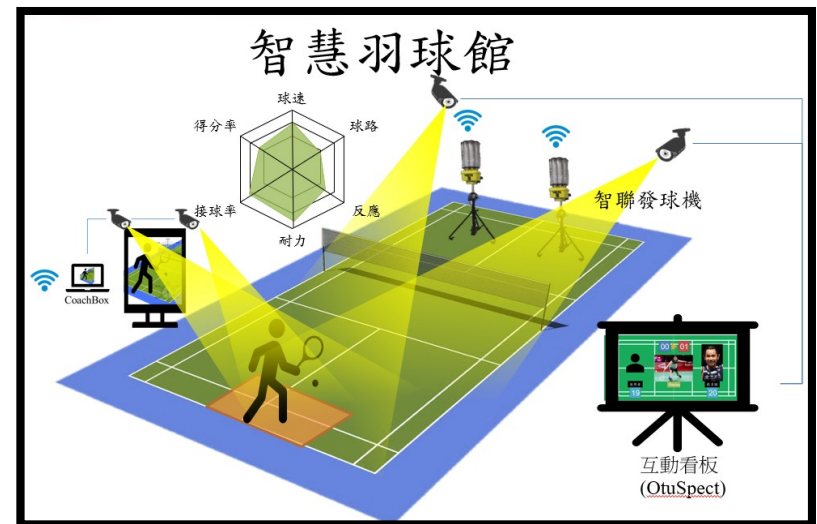
把所有道路狀況錄下後放入電腦中解析，由電腦自行理解每種狀況該有的駕駛行為，而非由專家將所有可能駕駛行為輸入電腦中。

何謂大數據分析？

➤ 大數據分析功用2：用找到的隱藏規則發展加值應用



透過辨識駕駛行為來判斷駕駛是否正在進行危險駕駛
(分類的應用)



把羽球選手隱藏的打球與移動習慣找出來，並實現

1. 糾正我們選手的動作與習慣
2. 找出對手的習慣與攻克方法
(迴歸預測的應用)

何謂大數據分析？

- 定義：廣義來說，大數據分析是由題目、資料與解題方法三部分組成的
- 題目：由廠商或是個人定義對廠商有利的題目
 - Key1：題目通常非常客製化
 - Key2：不同題目間差距一點點，解題方式可能差了十萬八千里
- 資料：所有與題目有關的資料都要被盡量收集
 - Key3：收集到的資料必須能涵蓋分析對象所有可能發生狀況
 - Key4：沒有資料就沒有辦法做大數據分析
 - Key5：魔鬼藏在細節中，資料是大數據分析的核心
- 解題方法：視題目定義與資料格式挑選合適方法
 - Key6：只要能達到目的，什麼方法都可以，不一定要特殊方法
 - Key7：以往都需要自己設計演算法，目前業界鼓勵使用套件



何謂大數據分析？

➤ 台灣地區大數據分析的成功案例

- 便利超商的熟食消費分析與櫃位配置
- 麥當勞報報
- 新光三越的櫃位配置
- 台積電的虛擬量測技術
- 鞋廠的成品瑕疵辨識



- 但其實不成功的分析案遠多餘成功的分析案。(請寫出以下案子在哪個環節出了問題)
- 案例1: XX醫院的”超大”數據分析
 - 案例2: 第一屆高速公路數據競賽的慘劇
 - 案例3: 某某花園的許願籤詩分析案
 - 案例4: 塑膠射出機廠商想做良率改善
 - 案例5: 國家太空中心的太陽能板分析案
 - 案例6: 資料分析廠商用深度學習模型開發出的刀具磨耗預測系統沒人要買



- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：專題題目思考

| 說在前頭：解決一個研究議題常見的方法

事前規劃

生產管理
設施規劃
作業研究
演算法

關鍵點：由專業知識與經驗設計方法

優點：不需要事先收集資料

缺點：常發生沒有考慮到的地方

小數據

事後分析

統計
品質管理
機器學習

大數據

大數據分析

關鍵點：不需要大量知識就可以設計方法

優點：只要數據夠大，就不會思考不周

缺點：需要事先收集資料與Labeling



國立中興大學

National Chung Hsing University

大數據分析 在產業上的應用

紅藍色代表不同群資料
實心點：被抽樣
空心點：未抽樣

事後分析

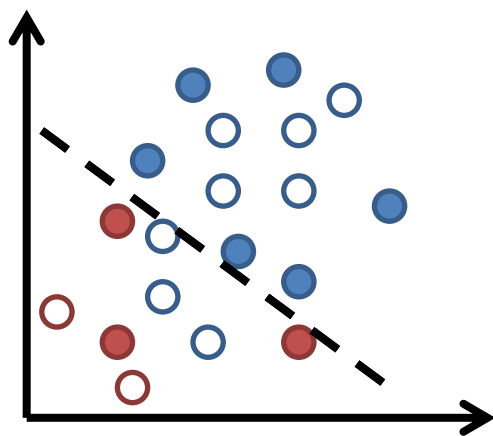
數據分析方法之作法說明

小數據

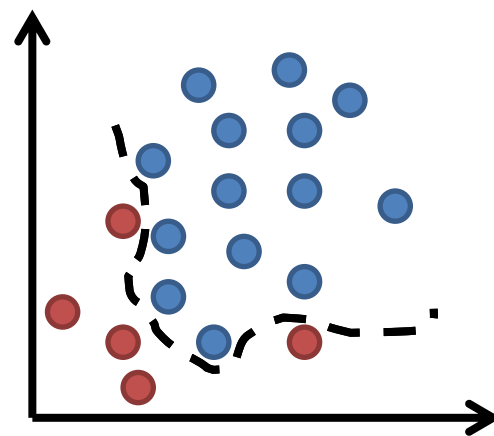
統計
品質管理
機器學習

大數據

大數據分析



小數據



大數據
(2016~)

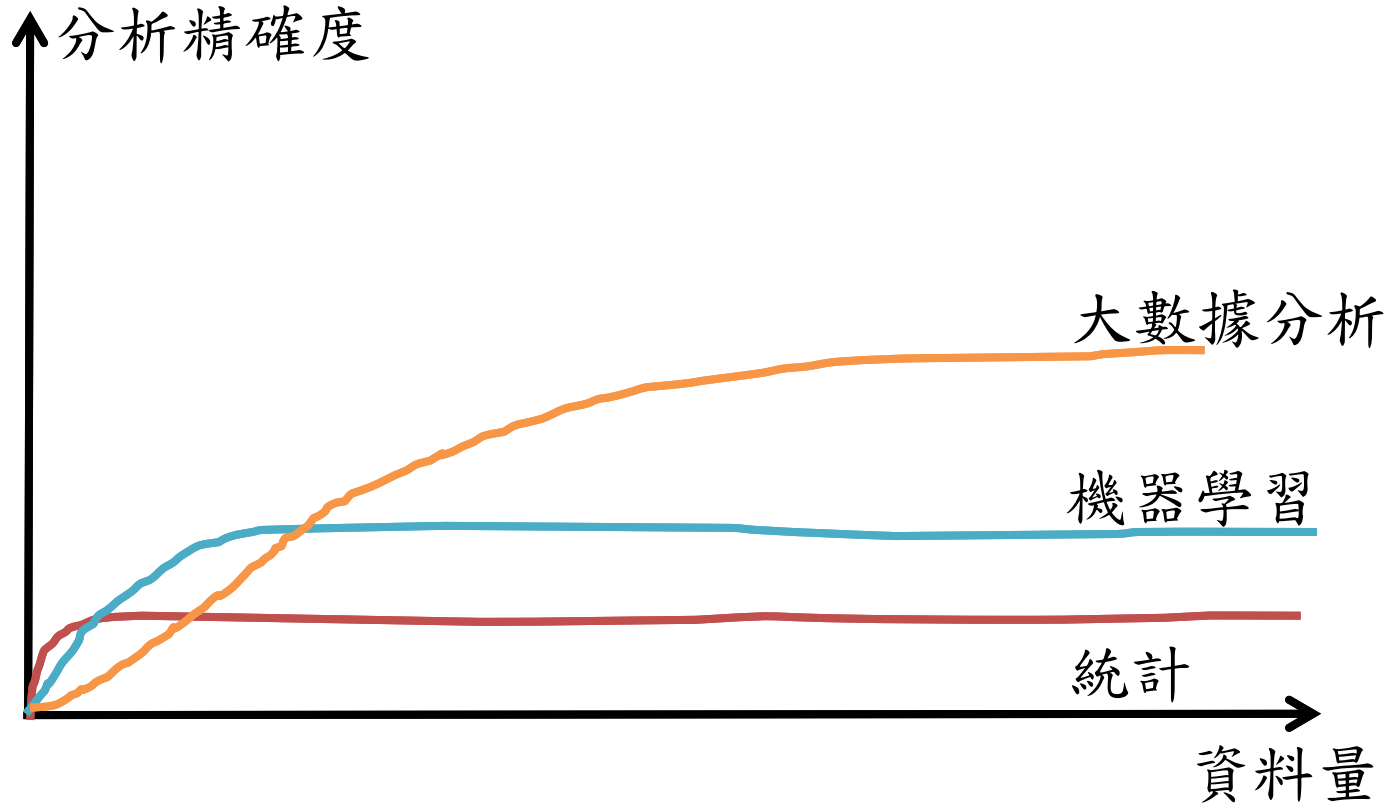
準確率

小 → 大

準確率



數據方法的效能分析



| 未來大數據分析這個領域的走向

事前規劃

生產管理
設施規劃
作業研究
演算法

由既有歷史資料
與大數據分析技術
重新建構專業知識

所謂的Explainable AI

小數據

大數據

事後分析

統計
品質管理
機器學習

大數據分析

由既有專業知識
加快數據分析流程
與分析準度



國立中興大學

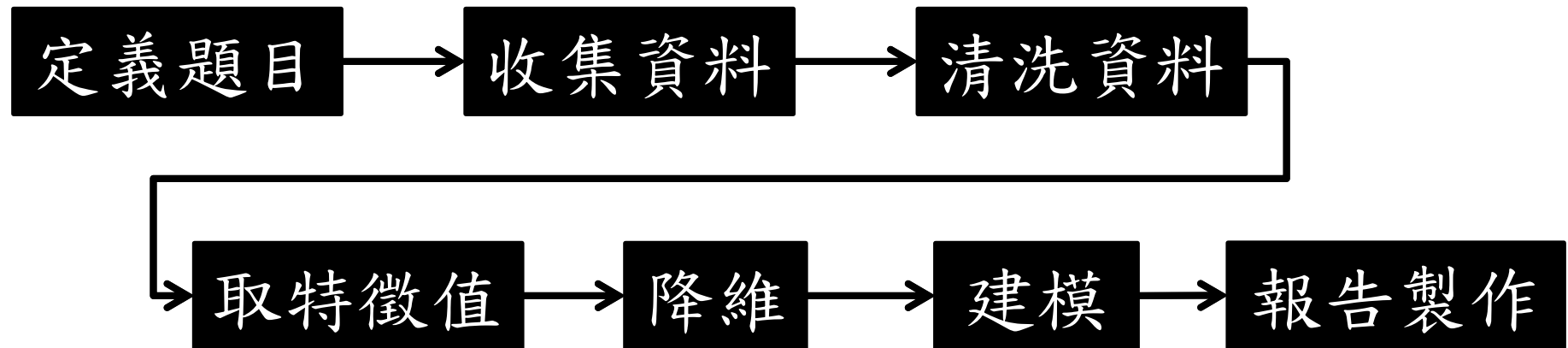
National Chung Hsing University

Outline



- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：專題題目思考

➤ 簡易分析流程



分類(活棋，可放在任何地方)

分析流程各部分定義

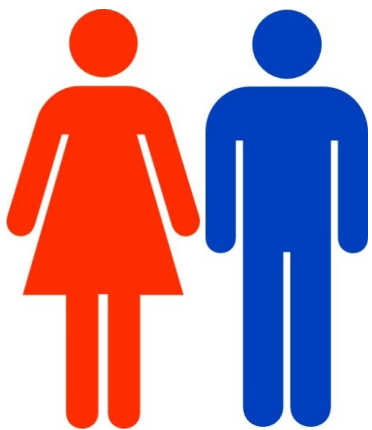


- **定義題目**：找出目標問題的輸入與輸出，並判斷解題類型(迴歸、預測、分類、分群、最佳化...)。
- **收集資料**：收集所有與本問題有關的資料集
- **清洗資料**：觀察資料、找出與修正不合理或缺值的資料、合併資料欄位
- **取特徵值**：(早期)由「人工」找出對資料分析有用的欄位
(近期)更改資料欄位型態以提昇辨識準確率
- **降維**：由「電腦」及「資料集」找出對資料分析有用欄位
- **建模**：給予輸入輸出後，建立起輸入輸出間的數學函數
- **分類**：將未知資料分到多個定義好的群組中
- **報告製作**：按照廠商的需求製作分析說明書



➤ 利用簡易案例加強分析流程的概念

- 案例1：將全班分成男生與女生



定義題目：定義目標案例實作環境

收集資料：對全班同學做問卷

清洗資料：去除或修正問卷內不合理的資料

取特徵值：由人從問卷中找出適合分辨的欄位

降維：由電腦與既有紀錄找出適合分辨的欄位

分類：利用既有資料建立分類模型

建模：不適用

分析及報告製作：撰寫執行SOP文件與解釋



國立中興大學

National Chung Hsing University

執行流程

➤ 利用簡易案例加強分析流程的概念

● 案例2：利用大數據分析降低學生餐廳廚餘量



定義題目：

收集資料：

清洗資料：

取特徵值：

降維：

分類：

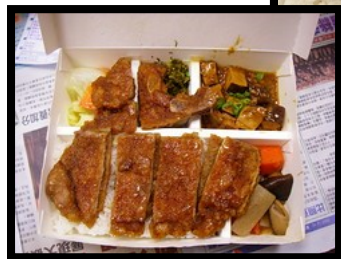
建模：

報告製作：



➤ 利用簡易案例加強分析流程的概念

- 案例2*：利用大數據分析降低學生餐廳廚餘量 (如何使用分類技術來提升分析準確度?)



定義題目：

收集資料：

清洗資料：

取特徵值：

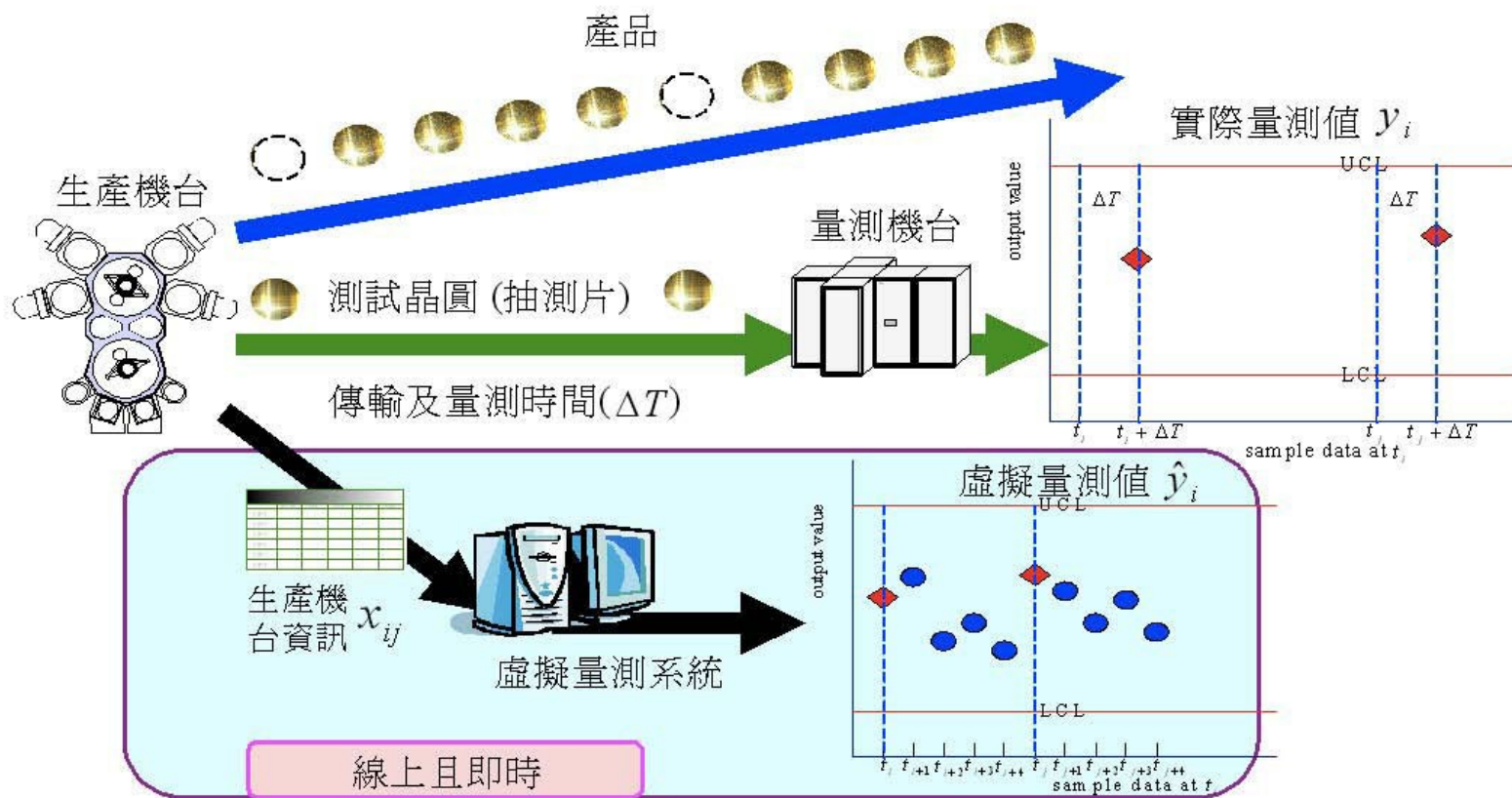
降維：

建模：

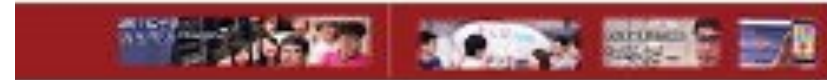
報告製作：



台積電的案例 V.S. 資料分析的流程



台積電的案例 V.S. 資料分析的流程



收集資料： 在機台上安裝許多感測器收集資料
(總共200個製程，每個製程收集200種資料)

清洗資料： 清除資料中不合理的資料

取特徵值： 找出40000種資料中，可能具有代表性者

降維： 找出所有特徵值中，對預測損壞最有效者

建模： 利用類神經網路、迴歸模型等建立模型

模型表示方式： $(y_1, y_2, \dots, y_n) = f(x_1, x_2, \dots, x_n)$

➤ 分析流程各部分執行關鍵(有緣人會需要)

● 定義題目

1. 沒有題目就不可能分析
2. 定義題目請從簡單到難
3. 與第一線人員而非主管對談

● 收集資料

1. 請確定資料是否可用，如視覺化，或了解資料收集方式
2. 請確定資料分佈跟題目所想是否一樣

● 清洗資料：

1. 大數據分析最關鍵地方
2. 技術不難，會if、for、while就會做
3. 但他是大數據分析最耗時、最痛苦的地方



➤ 分析流程各部分執行關鍵(有緣人會需要)

● 取特徵值

1. 盡可能把所有想到的都取出來，不要嫌麻煩

● 降維

1. 有很多套件可以用，請直接用套件

● 建模、分類：

1. 有很多套件可以用，初學者請直接用套件
2. 如果你想用大數據拿高薪，就要會改套件程式碼

● 報告製作：

1. 請知道廠商的終極目標
2. 針對廠商的終極目標製作成果報告

➤ 實際分析流程(通常分三輪實作)

● 第一輪（必然是失敗的分析，但可以當做分析基準）

1. 利用最簡易套件了解資料分佈
2. 利用最簡易套件了解題目的效能極限
3. 根據上面兩步驟設計第二三輪大方向

● 第二輪（失敗機率50%）

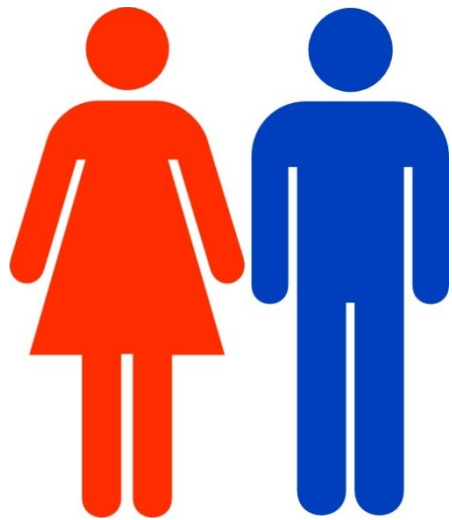
1. 依據上一輪失敗的經驗重新清洗資料
2. 利用大方向中的各項套件進行分析，並產生初步成果
3. 觀察結果是否還有需要改進的地方

● 第三輪（失敗機率80%）

1. 依據上一輪認定該改進的地方進行資料或演算法調整
2. 通常需要修改套件內的程式碼
3. 成功的分析，其目標是效能比分析基準好30%或符合廠商需求

➤ 利用簡易案例告訴你真實分析狀況...

● 案例1：將全班分成男生與女生



第一輪會遇到的狀況：

題目定義不恰當、男女比例失衡

有人亂填資料、預計方法效果很差

第二輪會遇到的狀況：

知道怎麼清洗資料、知道分析方向

但效果就是差一點

第三輪會遇到的狀況：

終於分析效能提昇，但也到交件時間

➤ 利用簡易案例告訴你真實分析狀況...

● 案例2：利用大數據分析降低雲科學生餐廳廚餘量



第一輪會遇到的狀況：

題目定義不恰當、資料收集錯誤或少收
餐廳資料不能量化、發生意想不到事件

第二輪會遇到的狀況：

跟餐廳員工確認題目細節、知道分析方向
定義好資料(如廚餘的定義)

第三輪會遇到的狀況：

終於分析效能提昇，但也到交件時間

➤ 利用簡易案例告訴你真實分析狀況...

● 案例3：台積電的案例



第一輪會遇到的狀況：

第二輪會遇到的狀況：

第三輪會遇到的狀況：

➤ 老師痛心疾首的附註說明...

大數據分析痛苦的都不是在技術端，在人，例如：

1. 分析對象自己要做啥都不知道

(題目無法定義)

2. 分析對象提供的資料無法分析

(無法量化、格式不對、資料嚴重傾斜)

3. 像變形蟲般的KPI

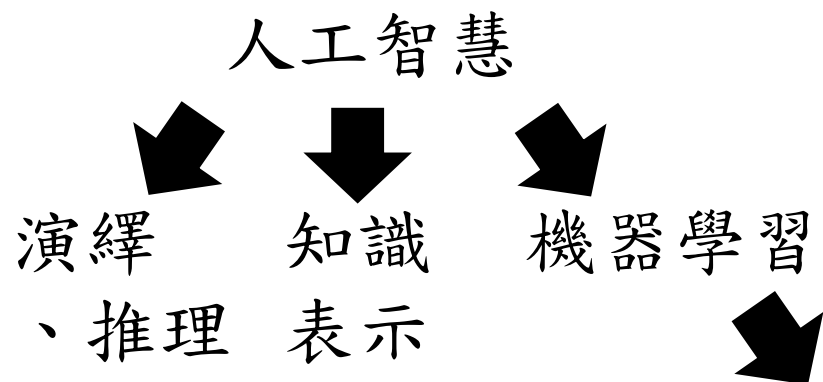
(本來說好做到70%準確度，下一次報告變80%)



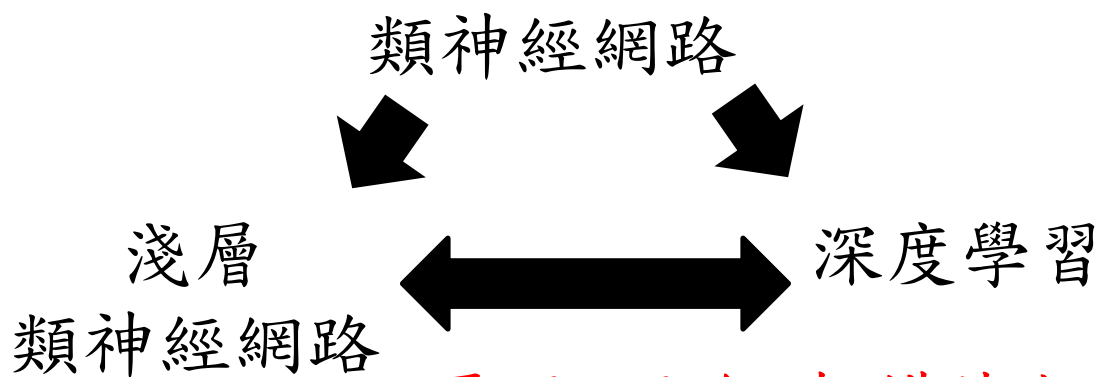
- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：專題題目思考



AI是什麼？



AI簡單來說就是用電腦
來學習輸入輸出間的規則



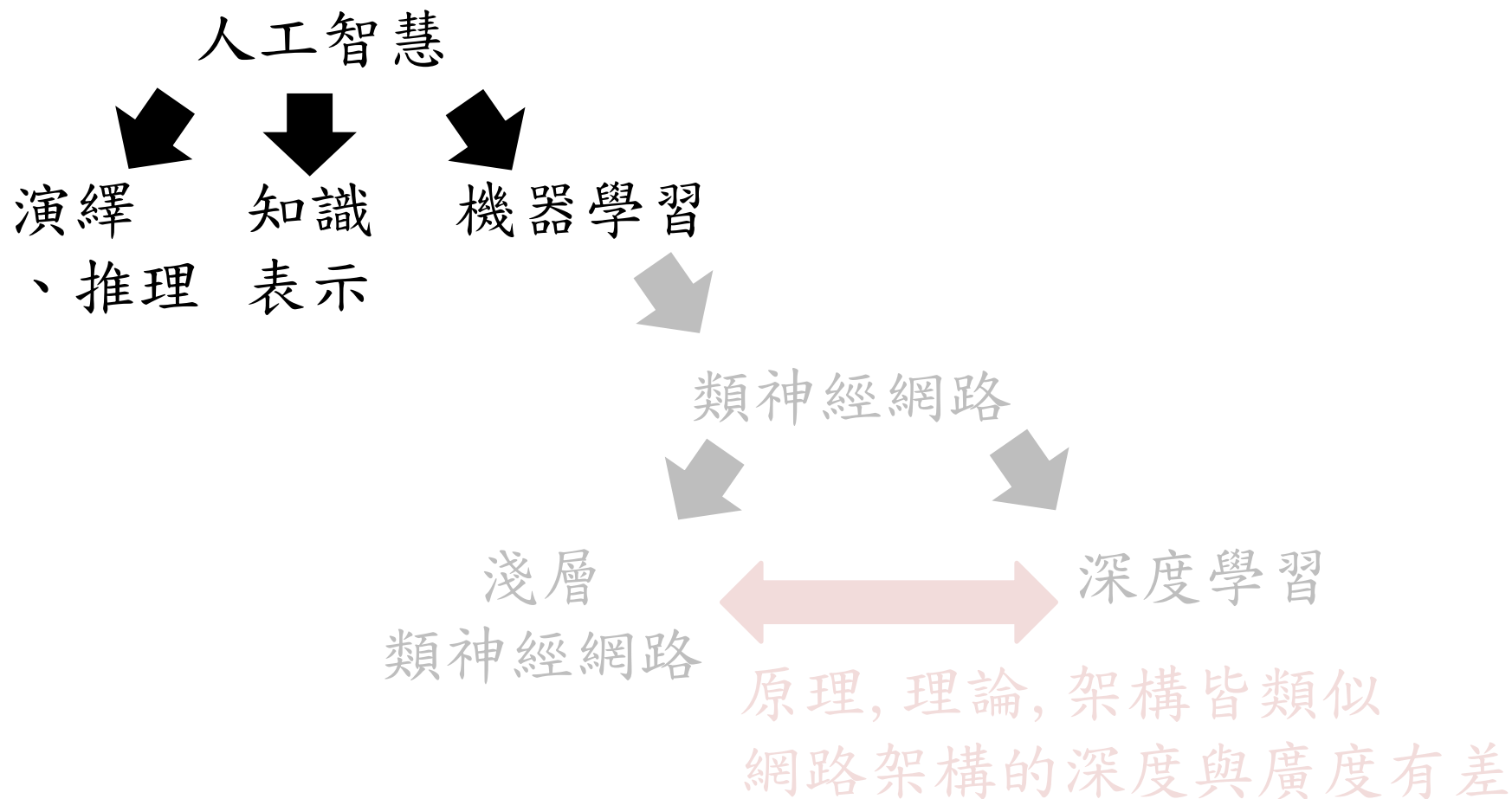
原理, 理論, 架構皆類似
網路架構的深度與廣度有差



國立中興大學

National Chung Hsing University

AI是什麼？



AI是什麼？

分支1：演繹、推理和解決問題

利用機率學直接模仿人類進行逐步的推理，就像是玩棋盤遊戲或進行邏輯推理時人類的思考模式。

請問O下一步要下哪裡？

O	X	X
X	O	
O		

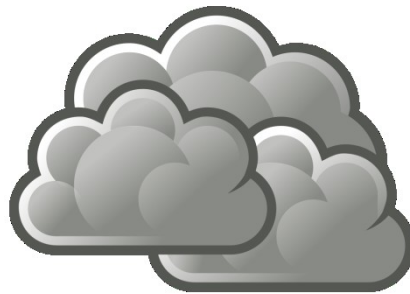
人類：用肉眼都知道右下角...

電腦：根據過往歷史紀錄，
下左邊有0%會贏
下下面有0%會贏
下右下角有100%會贏
所以我選右下角...

分支2：知識表示

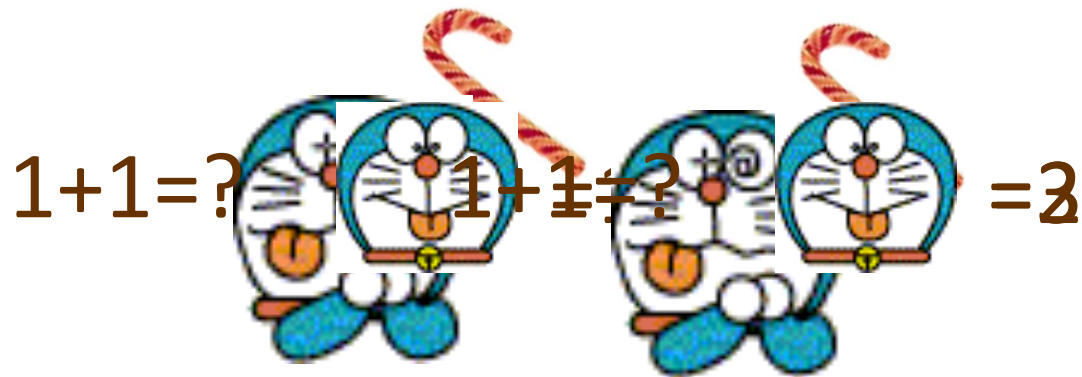
知識表示是人工智慧領域的核心研究問題之一，它的目標是讓機器儲存相應的知識，並且能夠按照某種規則推理演繹得到新的知識。

假設歷史紀錄告訴我們天氣只有晴天、陰天、雨天三種，且今天太陽不大且沒有下雨，請問今天天氣為何？



分支3：學習

主要目的是為了讓機器從使用者和輸入資料等處獲得知識，從而讓機器自動地去判斷和輸出相應的結果。

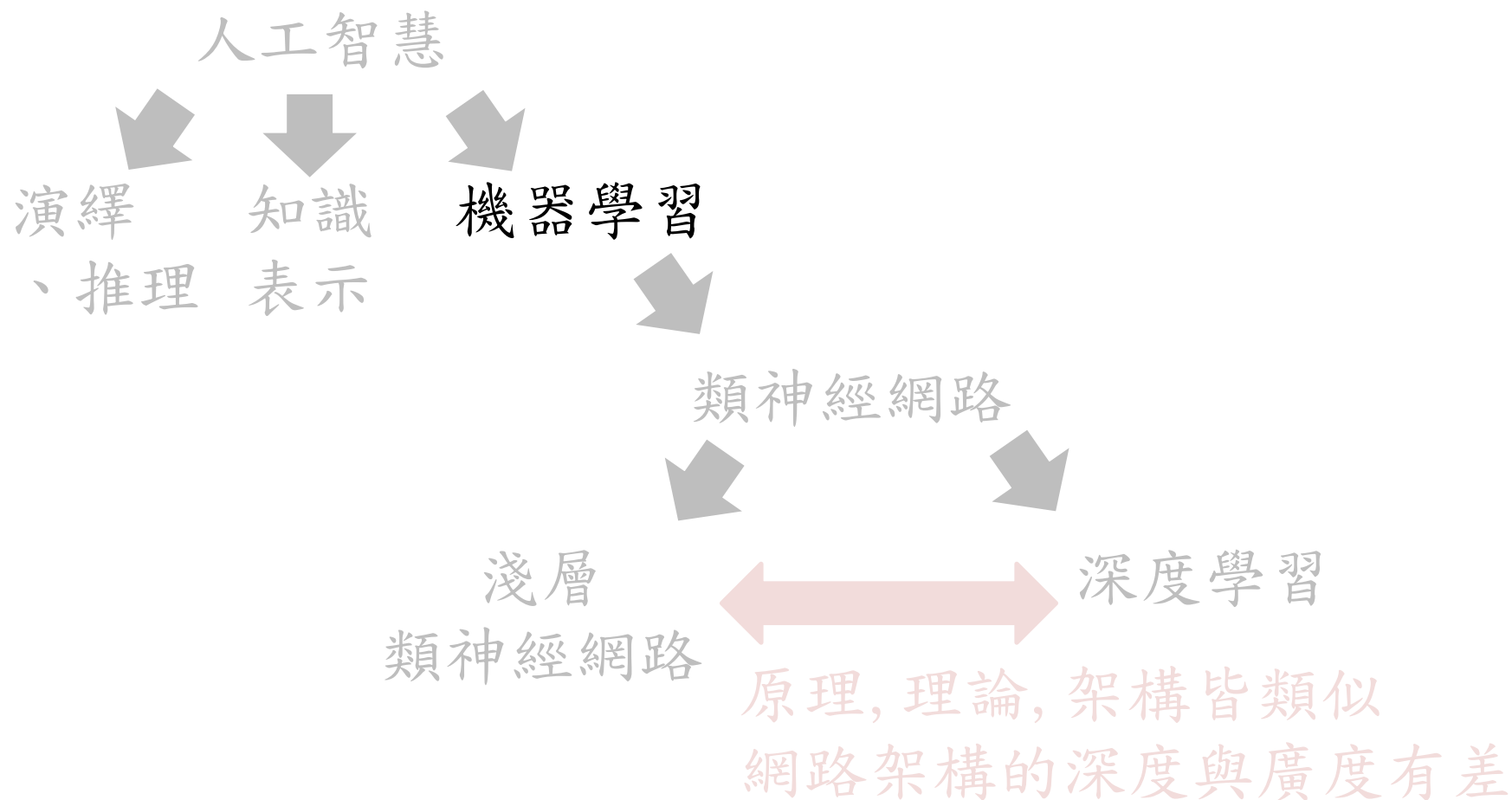




國立中興大學

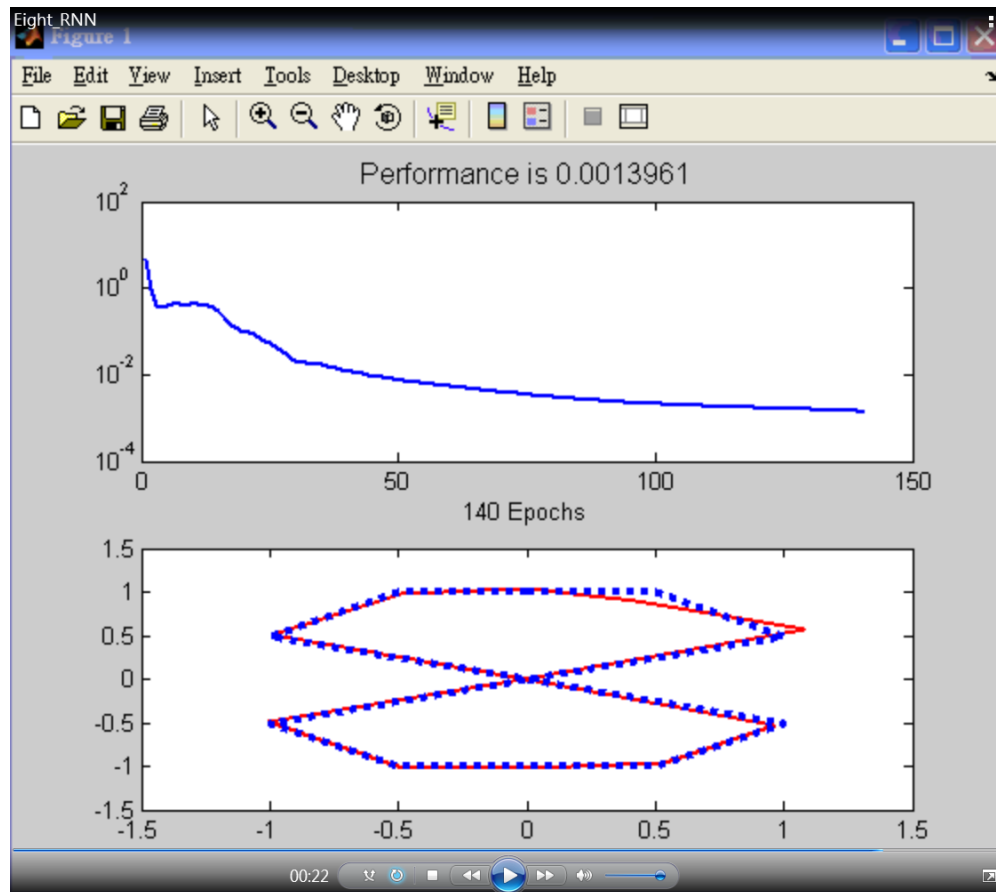
National Chung Hsing University

AI是什麼？



AI是什麼？

機器學習是對能通過經驗自動改進的電腦演算法的研究。
案例1: 教機器怎麼學會寫8



AI是什麼？

機器學習是對能通過經驗自動改進的電腦演算法的研究。
案例2: 利用機器學習校正大富翁的遊戲參數





國立中興大學

National Chung Hsing University

大富翁實作

問題背景

人數：3人

初始金額：5000元

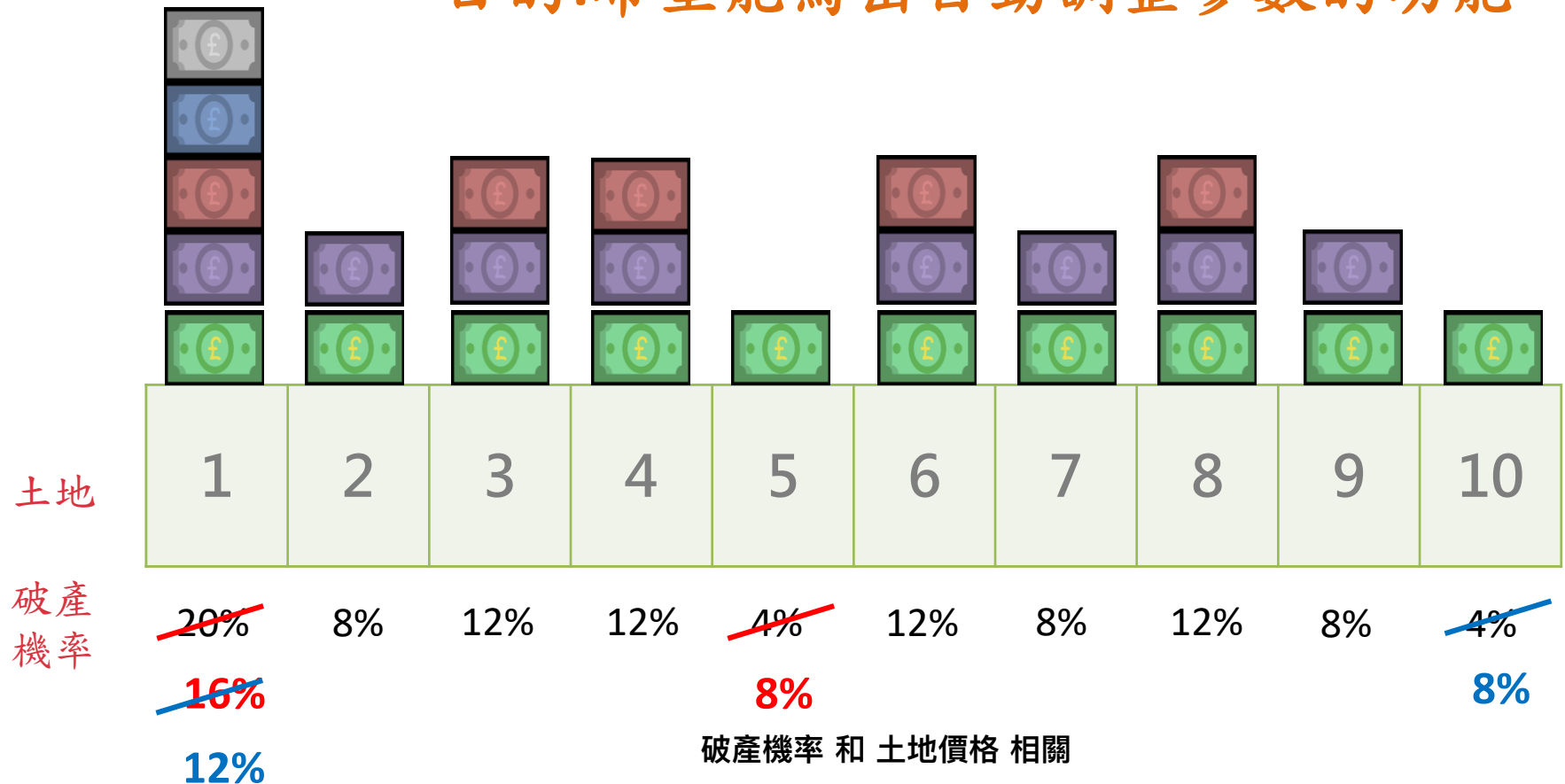
地圖大小：18格

功能：買土地、升級土地、收過路費

假設目前土地價格設錯，讓遊戲極度不平衡，
那我們該如何用AI來調整遊戲平衡呢？

解題概念

目的: 希望能寫出自動調整參數的功能



驗證問題合理性

在土地價格皆為500-700時，跑1000次看玩家在每點破產機率

looser_location	freq	rate
0	2	74
1	17	70
2	1	69
3	7	64
4	15	63
5	6	61
6	14	60
7	12	60
8	9	59

9	10	53
10	8	51
11	16	49
12	11	48
13	18	48
14	13	45
15	4	45
16		
17		

破產機率介於7.4%~3.9%左右

土地10價格改成2000時，跑1000次看玩家在每點破產機率

looser_location	freq	rate
0	10	909
1	1	612
2	14	595
3	16	588
4	17	570
5	3	552
6	11	552

9	15	534
10	6	524
11	18	523
12	12	522
13	8	521
14	7	493
15	5	487
16		
17		

證實價格打錯確實會造成破產機率提升

土地10破產機率變9.09%



驗證問題合理性

- 隨著局數增加，各點破產機率會接近平均分配。
- 選擇1000局作為下次調整地價的次數。
- 測試兩種調整方式：
 1. 每跑完1000局就讓破產率最高的地點金額下降100元
 2. 讓破產率最高的地點金額下降100元，破產率最低的地點提升100元



	looser_location	freq	rate
0	18	84	8.4
1	11	78	7.8
2	16	73	7.3
3	7	73	7.3

實作

調整方式1實作結果

- 42次調整
- 最高破產機率從 8.4% → 6.8%

	looser_location	freq	rate
0	1	68	6.8
1	8	68	6.8
2	14	64	6.4
3	7	63	6.3

土地價格表

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
land price	1040.0	800.0	210.0	1070.0	610.0	1170.0	1980.0	770.0	550.0	1700.0	1370.0	770.0	1000.0	1190.0	450.0	1400.0	410.0	1920.0
level up	312.0	240.0	63.0	321.0	183.0	351.0	594.0	291.0	165.0	510.0	411.0	231.0	300.0	357.0	135.0	420.0	123.0	576.0
pass fee	520.0	400.0	105.0	535.0	305.0	585.0	990.0	385.0	275.0	850.0	685.0	385.0	500.0	595.0	225.0	700.0	205.0	960.0



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
land price	940.0	800.0	210.0	970.0	610.0	970.0	1380.0	870.0	550.0	900.0	1370.0	670.0	900.0	890.0	450.0	900.0	410.0	1020.0
level up	282.0	240.0	63.0	291.0	183.0	291.0	414.0	261.0	165.0	270.0	261.0	201.0	270.0	267.0	135.0	270.0	123.0	306.0
pass fee	470.0	400.0	105.0	485.0	305.0	485.0	690.0	435.0	275.0	450.0	435.0	335.0	450.0	445.0	225.0	450.0	205.0	510.0



	looser_location	freq	rate
0	5	91	9.1
1	6	81	8.1
2	4	78	7.8
3	12	69	6.9

翁實作

調整方式2實作結果

- 16次調整
- 最高破產機率從
9.1% → 6.9%

土地價格表

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
land price	950.0	1810.0	820.0	1830.0	1890.0	1930.0	600.0	1200.0	650.0	460.0	370.0	1530.0	680.0	580.0	1580.0	580.0	470.0	880.0
level up	285.0	543.0	246.0	549.0	567.0	579.0	180.0	360.0	195.0	138.0	111.0	459.0	204.0	204.0	474.0	174.0	141.0	264.0
pass fee	475.0	905.0	410.0	915.0	945.0	965.0	300.0	600.0	325.0	230.0	185.0	765.0	340.0	340.0	790.0	290.0	235.0	440.0



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
land price	950.0	1610.0	1730.0	1830.0	1590.0	1430.0	1310.0	1730.0	1710.0	1790.0	1930.0	1530.0	1790.0	1830.0	1630.0	1530.0	1710.0	1630.0
level up	285.0	483.0	489.0	519.0	477.0	429.0	513.0	489.0	483.0	507.0	549.0	459.0	537.0	519.0	459.0	429.0	513.0	459.0
pass fee	475.0	805.0	815.0	865.0	795.0	715.0	855.0	815.0	805.0	845.0	915.0	765.0	895.0	865.0	765.0	715.0	855.0	765.0

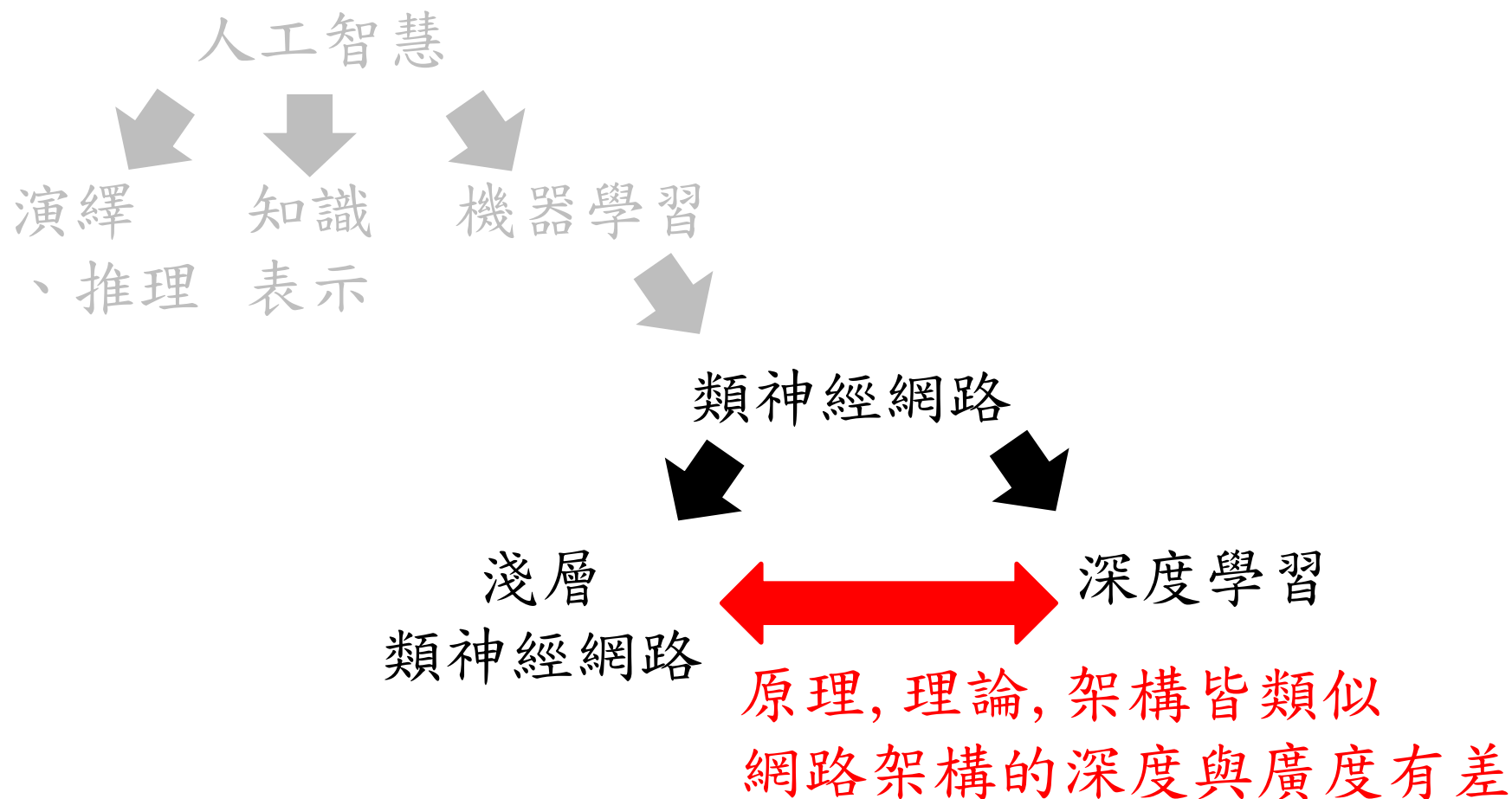
	looser_location	freq	rate
0	3	69	6.9
1	4	68	6.8
2	8	67	6.7
3	6	65	6.5



國立中興大學

National Chung Hsing University

AI是什麼？



類神經網路？

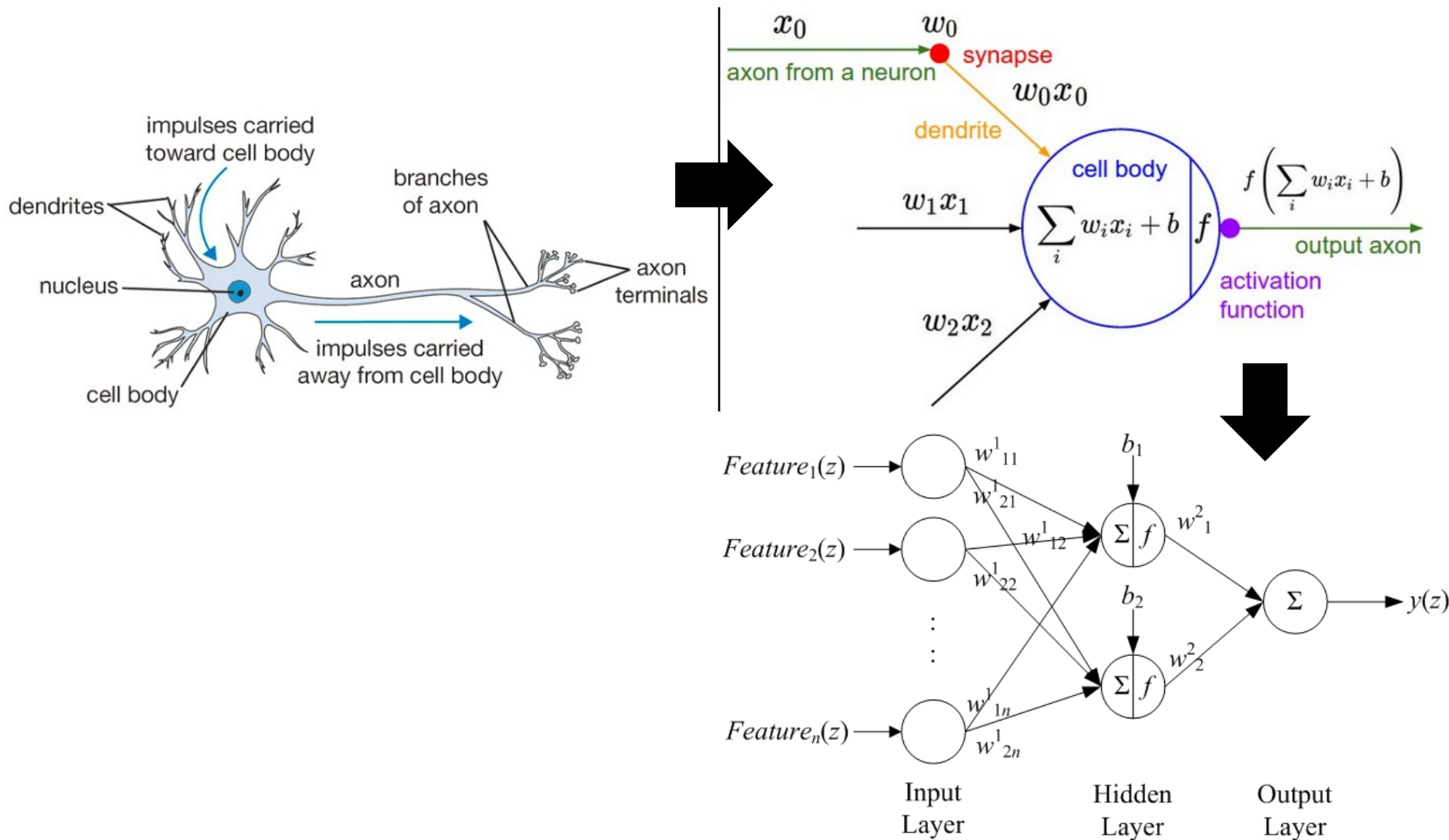
概念一：他是模擬人類的神經元所設計而成的。

概念二：他具有學習的功能，其學習的流程就像你做完考卷後，會利用訂正的過程來讓自己學會考題上的知識

概念三：他是個黑盒子，你不需要了解他的內部操作就可以使用他

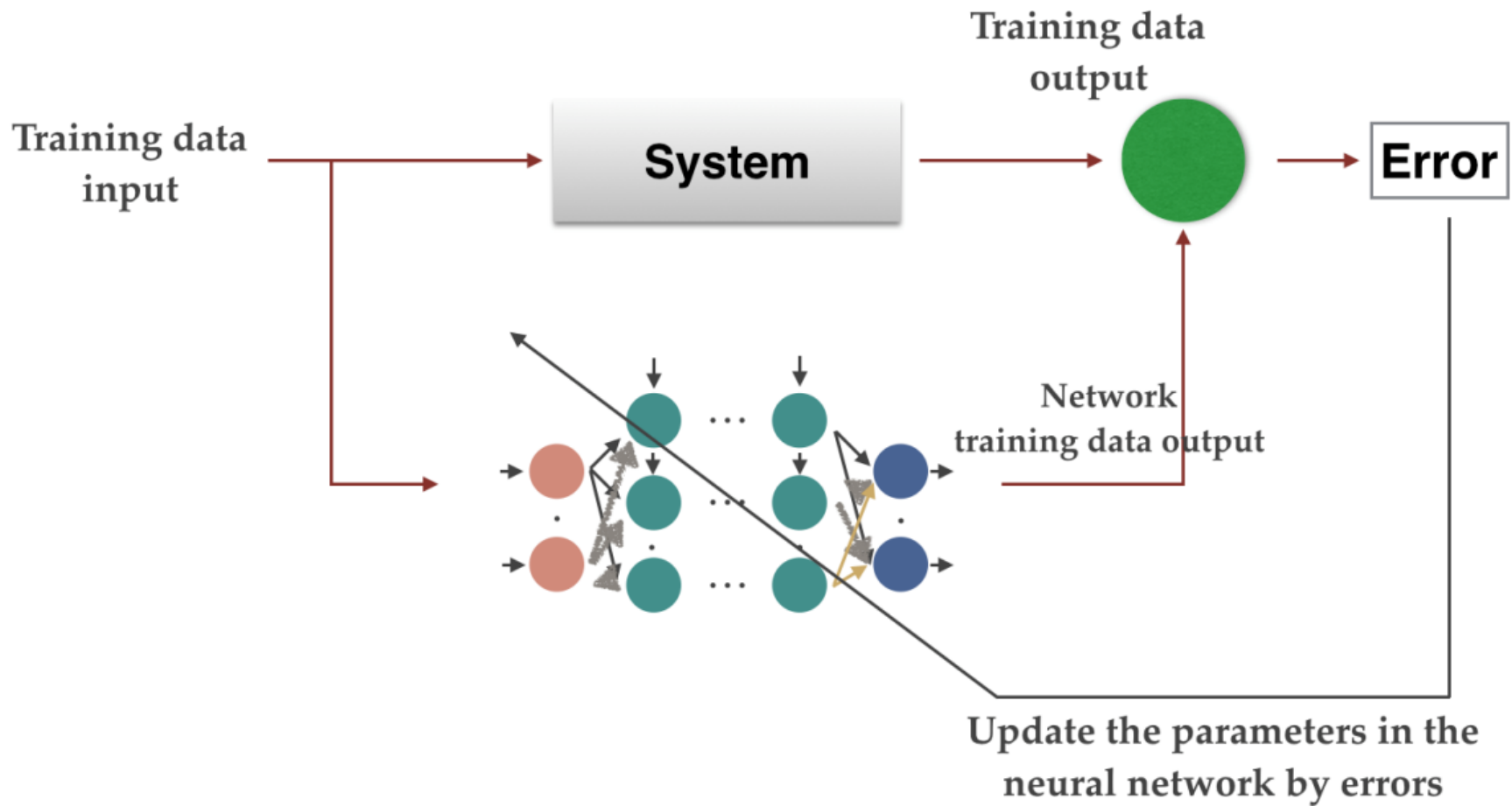


類神經概念一：他是模擬人類的神經元所設計而成的



AI是什麼？

類神經概念二：他具有學習的功能，其學習的流程就像你做完考卷後，會利用訂正的過程來讓自己學會考題上的知識

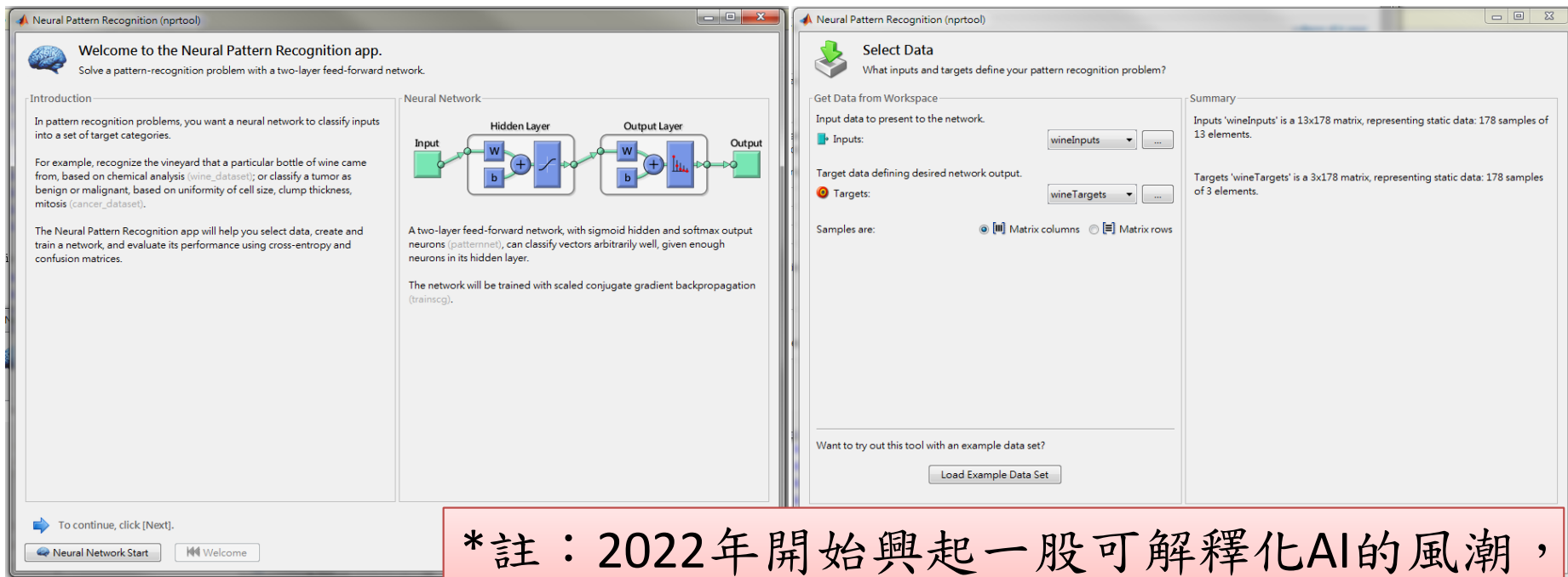


類神經概念三：他是個黑盒子，你不需要了解他的內部操作就可以使用他

2010年以前：要使用類神經網路，就必須了解類神經網路的架構，自己用程式刻出來。

2010年以後：幾乎所有主要軟體都有類神經網路的套件，你只要在套件中input 輸入輸出後，就可以快速建立類神經網路了

類神經概念三：他是個黑盒子，你不需要了解他的內部操作就可以使用他



Neural Pattern Recognition (nprtool)

Welcome to the Neural Pattern Recognition app.
Solve a pattern-recognition problem with a two-layer feed-forward network.

Introduction

In pattern recognition problems, you want a neural network to classify inputs into a set of target categories.

For example, recognize the vineyard that a particular bottle of wine came from, based on chemical analysis (`wine_dataset`); or classify a tumor as benign or malignant, based on uniformity of cell size, clump thickness, mitosis (`cancer_dataset`).

The Neural Pattern Recognition app will help you select data, create and train a network, and evaluate its performance using cross-entropy and confusion matrices.

Neural Network

Input → Hidden Layer → Output Layer → Output

A two-layer feed-forward network, with sigmoid hidden and softmax output neurons (`patternnet`), can classify vectors arbitrarily well, given enough neurons in its hidden layer.

The network will be trained with scaled conjugate gradient backpropagation (`trainscg`).

Select Data

What inputs and targets define your pattern recognition problem?

Get Data from Workspace

Input data to present to the network.

Inputs: `wineInputs`

Target data defining desired network output.

Targets: `wineTargets`

Samples are: ☒ Matrix columns ☐ Matrix rows

Summary

Inputs '`wineInputs`' is a 13x178 matrix, representing static data: 178 samples of 13 elements.

Targets '`wineTargets`' is a 3x178 matrix, representing static data: 178 samples of 3 elements.

Want to try out this tool with an example data set?

Load Example Data Set

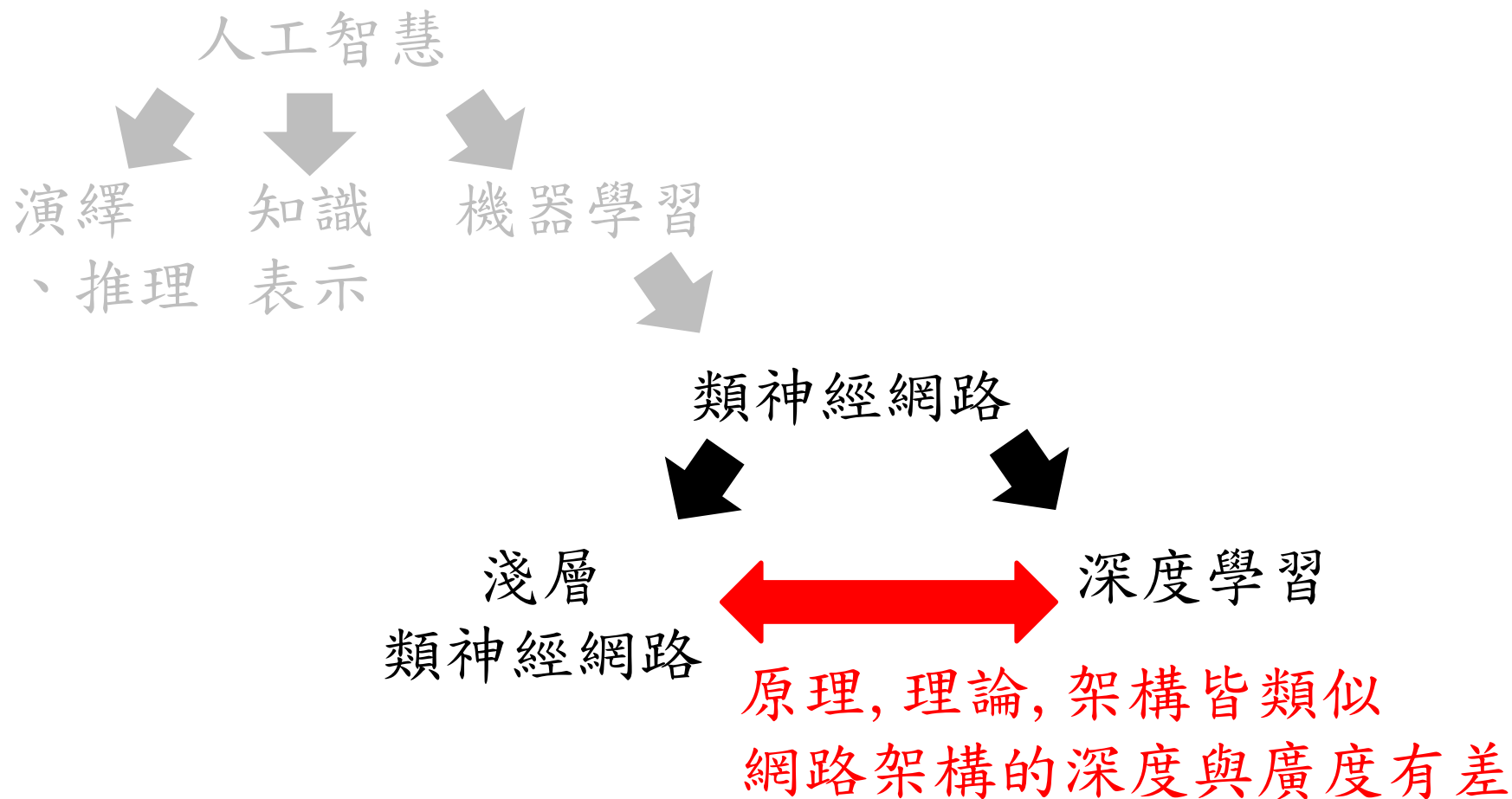
*註：2022年開始興起一股可解釋化AI的風潮，因此了解AI內部核心原理變得越來越重要



國立中興大學

National Chung Hsing University

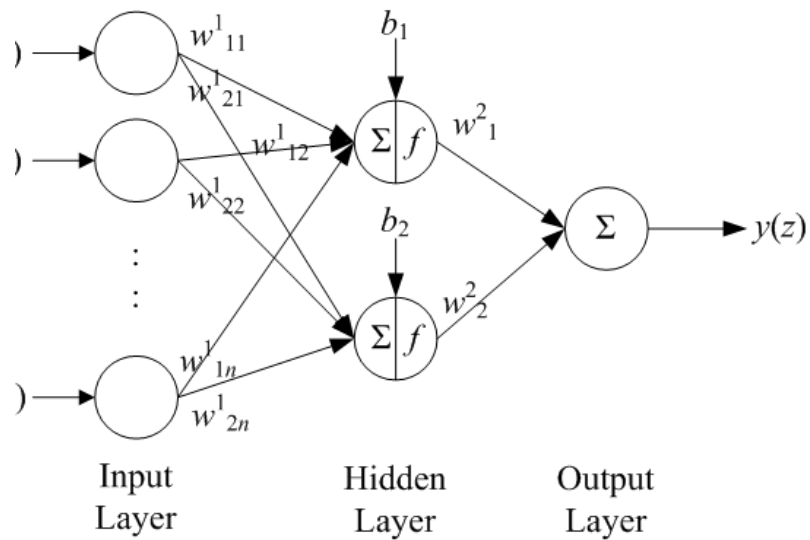
AI是什麼？



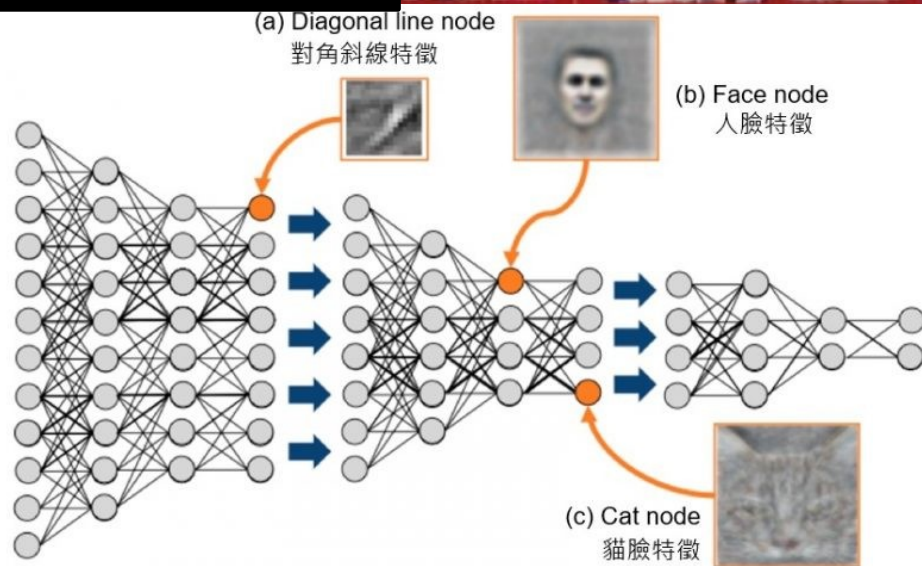


AI是什麼？

淺層類神經網路：



深度學習：



深度學習講得實在一些，其實就是比較多層的淺層類神經網路，理論、方法都一樣。

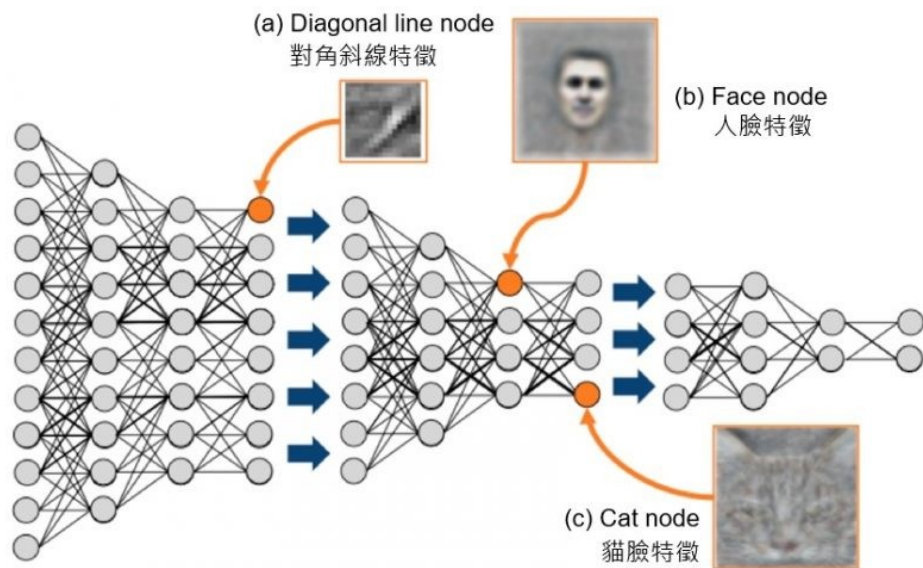
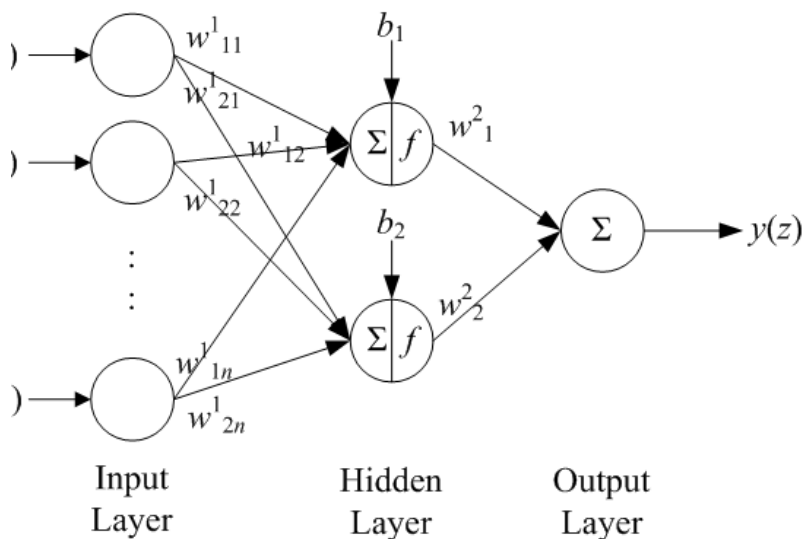
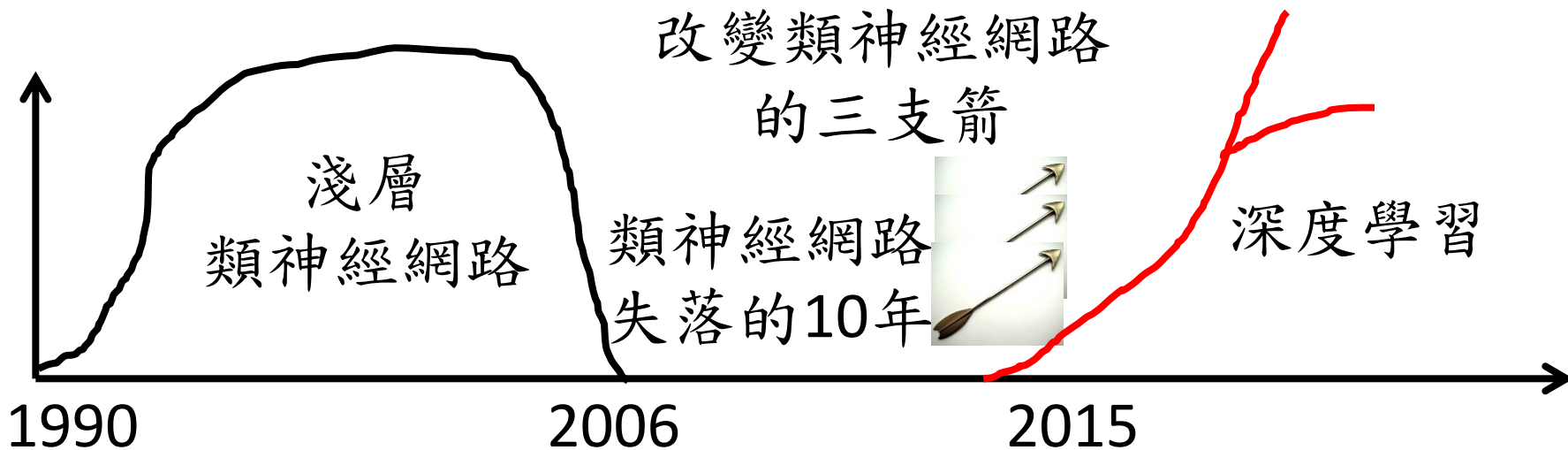
那為什麼需要深度學習呢？



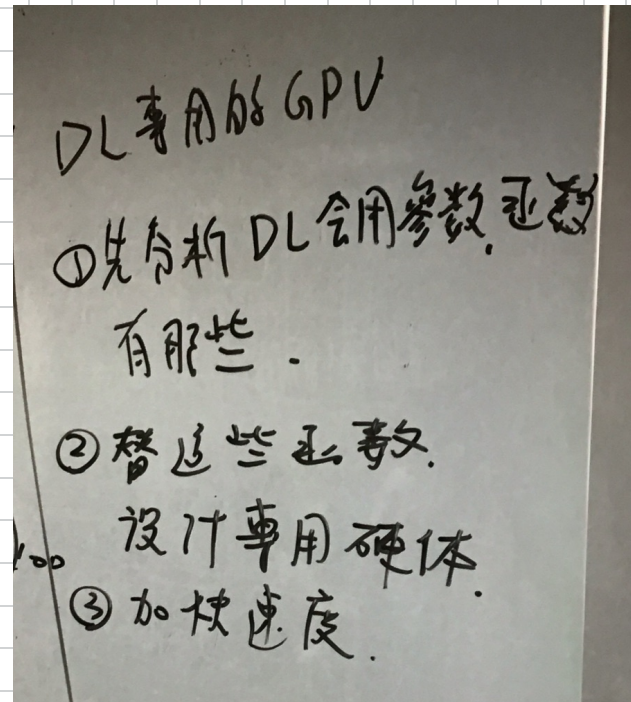
國立中興大學

National Chung Hsing University

AI是什麼？



某些函數 很常用 就會被 直接做成硬件

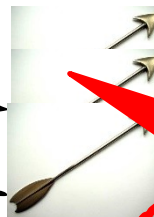
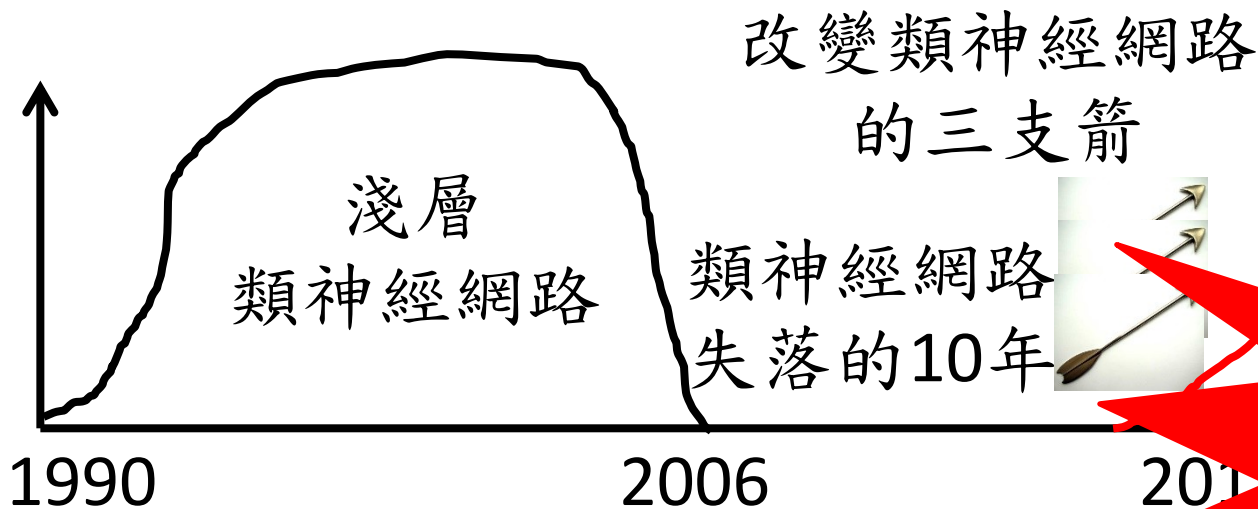




國立中興大學

National Chung Hsing University

AI是什麼？



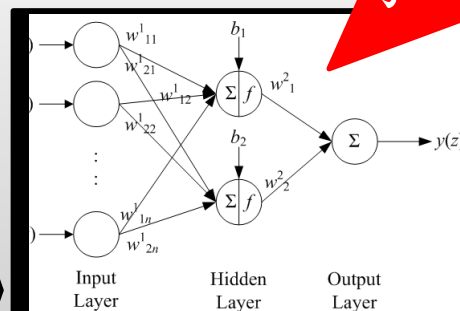
深度學習

類神經網路核
心需要自己寫，
跨入門檻高

淺層類神經網路之介紹

資料

取特徵值, 分類,
降維等資料前處
理演算法

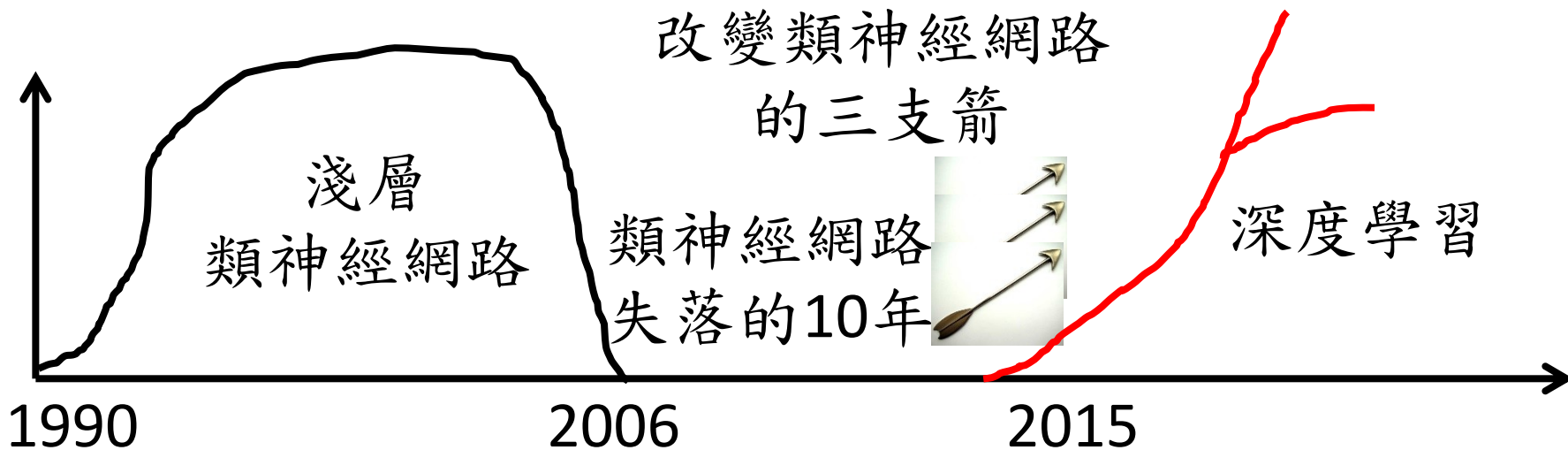


自行設計
類神經網路

結果



AI是什麼？



類神經網路失落的10年

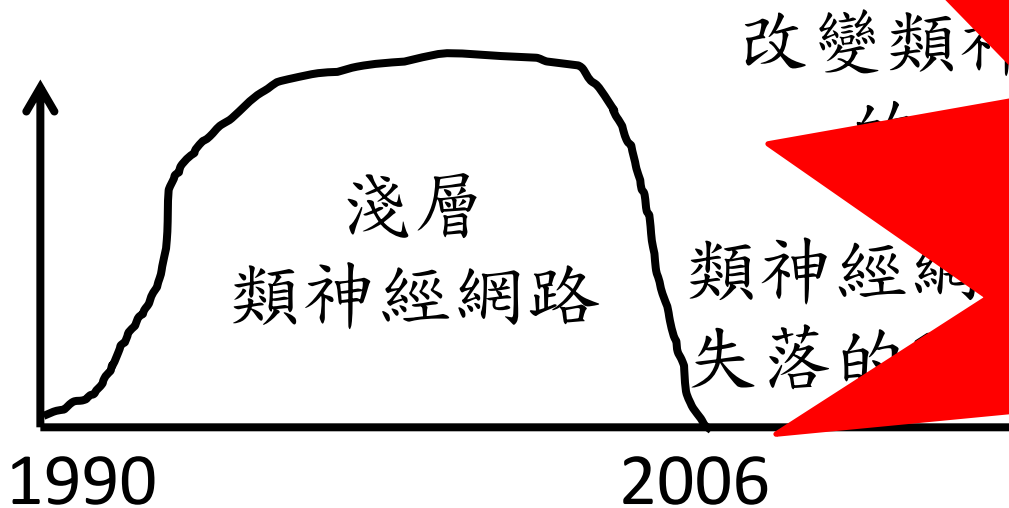
1. 類神經網路參數有限,解決問題能力有限(90~95%)
2. 知道增加類神經網路的參數可以增加效能,但硬體能力不足
3. 淺層類神經網路的效能受限於資料前處理結果的好壞,相關論文最後都變成討論資料前處理演算法



國立中興大學

National Chung Hsing University

AI是 麼？



大家都會用深度學習模型，但是不懂怎麼用，怎麼調，怎麼從裡面挖掘有用資訊

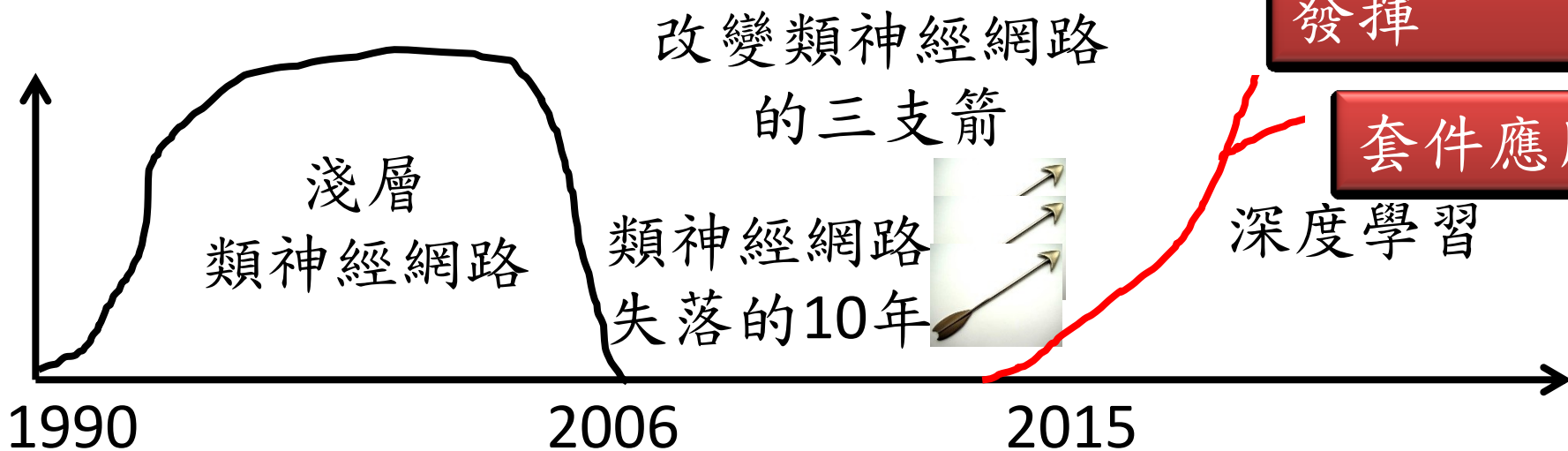
改變類神經網路的三支箭

1. 雲端計算(門檻高)→硬體效能增加,可以訓練更大的類神經,所以深度學習的概念出現→Alpha Go
2. GPU→可用單機訓練深度學習模型→降低使用門檻
3. 程式套件化的概念出現→深度學習模型套件出現→再度降低使用門檻



核心價值
發揮

套件應用

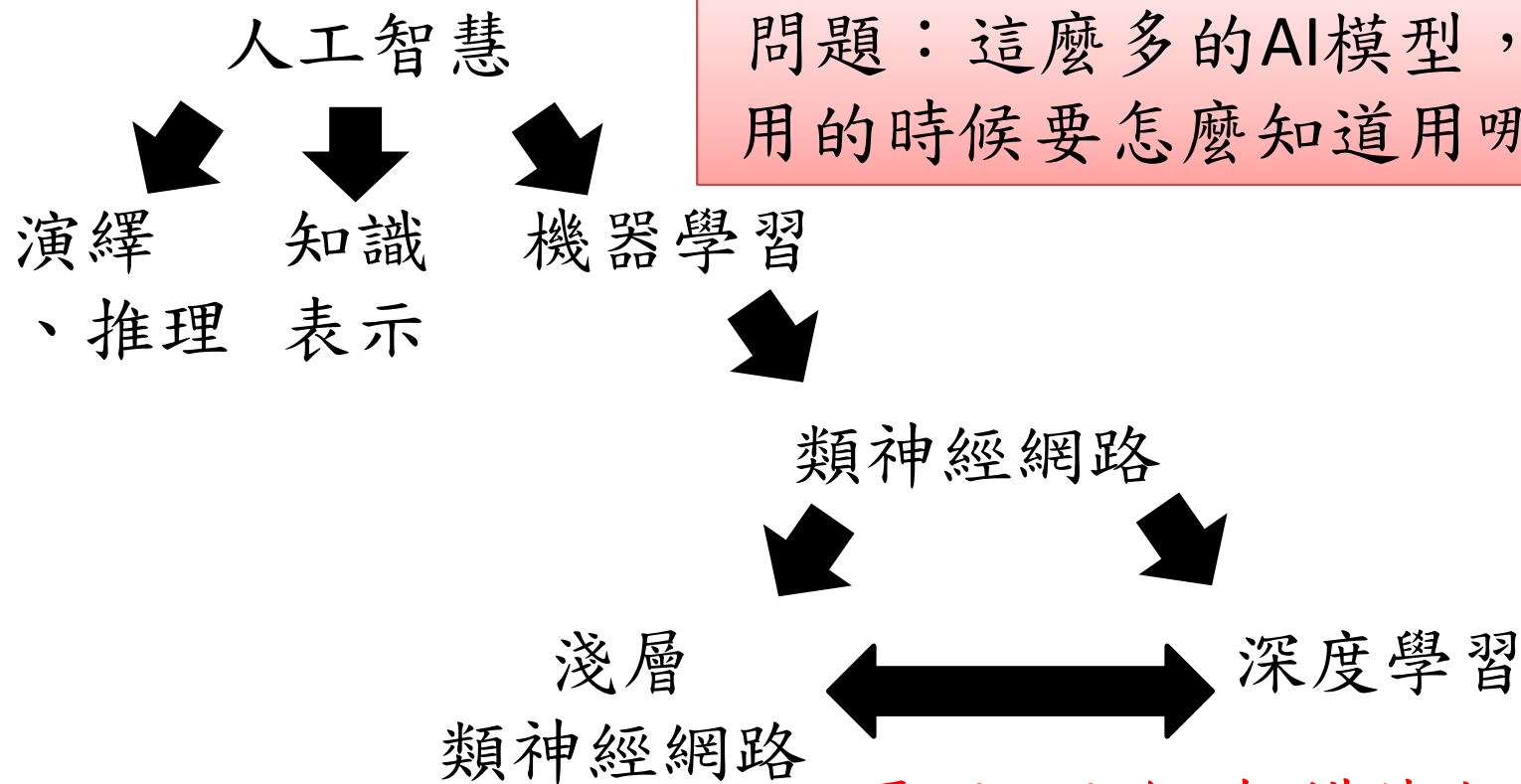


深度學習的分支

1. 套件應用: 用套件解決各項問題, 價值低, 成本高
(一定得有GPU才能將訓練好的模型套在各種應用中)
2. 核心價值發揮: 從訓練好的深度學習模型挖掘各種資訊, 提高分析效能及降低使用成本



AI是什麼？



原理, 理論, 架構皆類似
網路架構的深度與廣度有差

► 依照你題目與資料類型挑選合適方法

資料類型 處理方式	離散資料	時序資料	...	時空資料
降維	LDA、PCA	FNN	...	-
分群分類	KNN、 K-means	K-medoids...	...	DBSCAN
:	:	:	...	:
建模	DNN	RNN、LSTM...	...	CNN+LSTM



國立中興大學

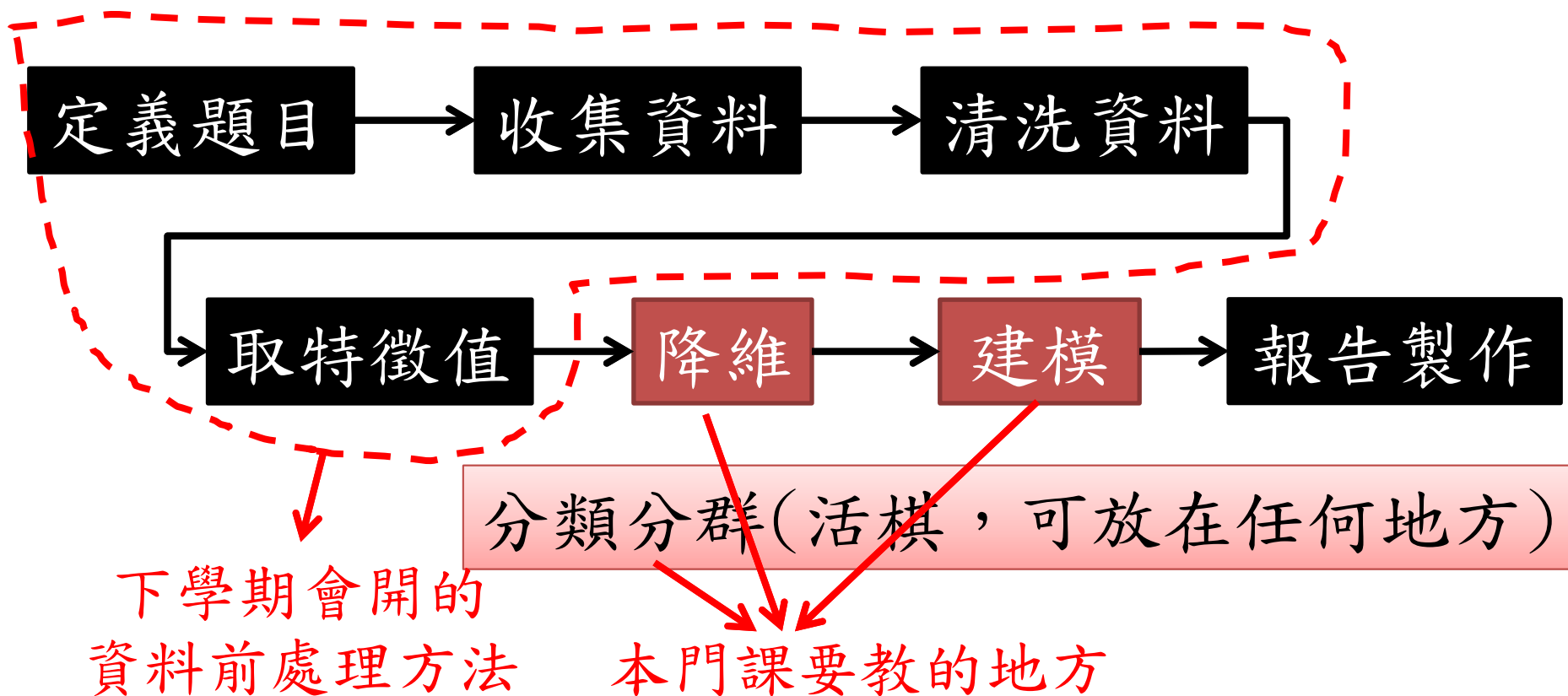
National Chung Hsing University

Outline

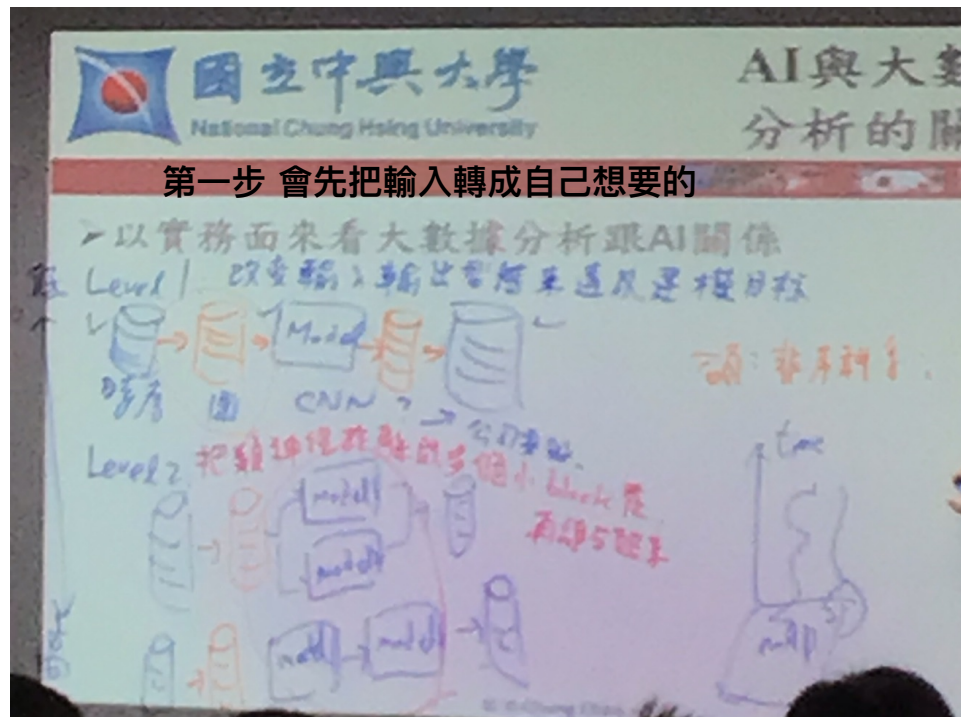


- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：專題題目思考

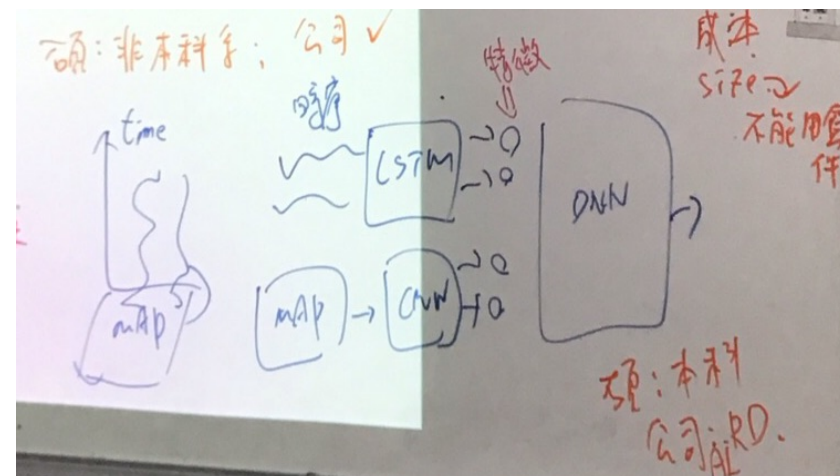
➤ 以純理論觀點來看大數據分析跟AI關係



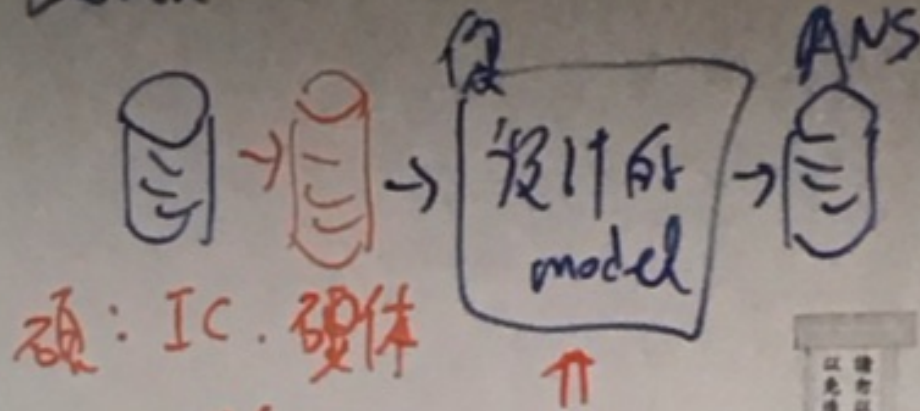
➤以實務面來看大數據分析跟AI關係



會把模組 裡面的東西 拆成小的神經元
自己組合 在做成新的模組



Level 3.



公司. IC.
ALRD

符合. 环境
限制.
成本.

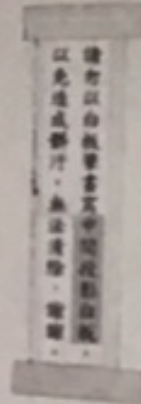
size

不能用套
件.

特徵
↓

TM
~O
~O

DNN





AI與大數據 分析的關係？

來看看你人生做資料分析的狀況：

	資料前處理	使用模型
大學	已整理好的乾淨資料 (如Kaggle、廠商資料)	套模仔
碩士	已整理好的乾淨資料 (如Kaggle、廠商資料)	套模仔或自行開發
就業	非常骯髒的資料 (現場的第一手資料)	套模仔 (公司審核過的模型)



This is why AI talents trained in schools are not what the industry needs



AI與大數據 分析的關係?

為什麼本門課要特別把前處理部分獨立出來教?

所以你出去工作時的真實情況:

	資料前處理	使用模型
處理時間	超過1/2的計畫時間	頂多1/2的計畫時間
遇到問題?	沒有正解，靠經驗	都套模，很難出錯 (如果會改模型薪水高)

結論1：不用程式強也可成為好的資料分析師，因為資料前處理不用太多程式技巧，要的是經驗。

結論2：如果你要在資料分析師這個行業拿到高薪，你必須(1)很有處理資料的經驗，(2)會改模型套件



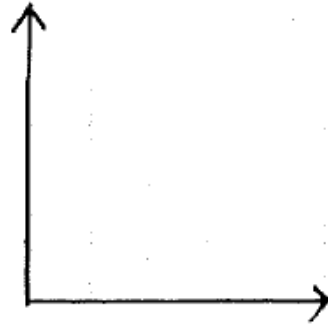
- 大數據分析概念介紹
- 大數據分析在產業中的定位
- 執行大數據分析所需的流程
- AI是什麼？
- 大數據分析與AI間的關係
- 課程活動1：本日課程快速複習
- 課程活動2：找出工業或商業大數據案例各一則
- 課程活動3：專題題目思考



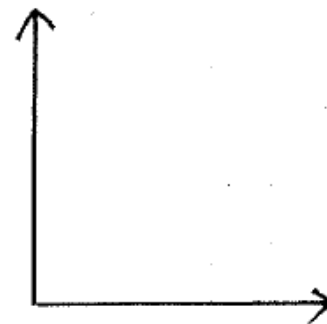
課程活動1

1. 請寫出大數據分析為何效果比小數據方法好的原因：

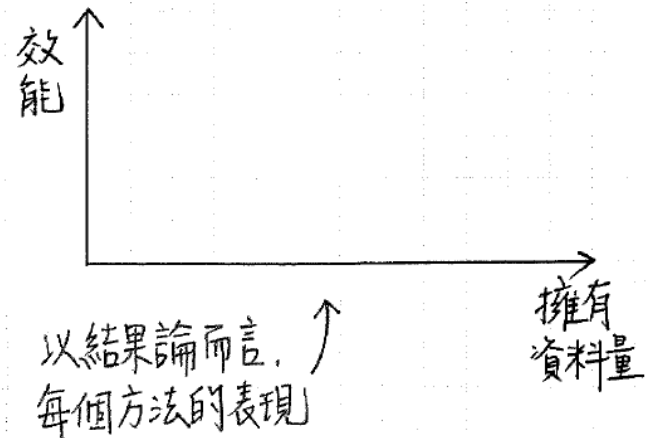
統計(小數據)



大數據



2. 請寫出統計方法、機器學習，還有深度學習在不同資料量下的效能。



這裡的課程活動 會考



3. 請寫出講義所提及「利用大數據分析降低學生餐廳廚餘量」分析案之每個分析流程要做什麼？
4. 請寫出「台積電虛擬量測技術」在每一輪分析時會遇到什麼困難



國立中興大學

National Chung Hsing University

課程活動2



1. 目前台灣已有許多大數據案例被公布出來，請試著上網找尋工業與商業各一個案例，並簡略說明他們的「題目」、「資料集」為何？

1. 請每組提出想做的大數據分析題目10個
2. 與老師討論每個題目是否適合
3. 在老師挑選出其中最適合者後，寫出其分析流程