

和AI做朋友

相識篇

開啟AI任意門



中小學人工智慧教育教學示範例系列

國中階段

李建樹 · 編 | 陳虹伶、陳瑞翎 · 執行編輯 | 陳建中 · 吉佛慈 · 蔡淑芬 · 盧佳君 · 合著

序言

早在AlphaGo打敗圍棋棋王以前，「AI」已經逐漸進入我們生活周遭，例如手機上的語音助理、自駕車，甚至是以演算法預測顧客的喜好以提供更適切的商品，AlphaGo不過是這場AI大戲中較引人矚目的一個角色。

當AI從研究室走入產業界、走入你我生活，成為不可忽視的存在，教育部認為不該只著重高階研發人才的培育，也應該向下扎根。讓中小學生有機會體驗 AI、知道 AI 的應用與對自己未來及生活的影響；並針對AI 原理及技術有興趣的學生，提供進階學習的資源及管道。

基於這樣的理念，我們邀請計畫團隊依十二年國民基本教育課程綱要、教師執教經驗與中小學學生相關數理觀念建立情形，編纂《和AI做朋友》這套教材。編輯團隊依據學生回饋修正教材，也將這些內容編輯成教案，做為其他教師後續實施參考的依據。

《和AI做朋友》是教育部為有興趣學習與教授AI的學生及教師鋪墊的第一步，是想一窺AI奧秘的師生們的參考教材。後續將陸續發展不同主題的課程內容，以及相對應的數位課程、實作課程與融入式課程等等，以提供師生們更多元的學習需求。

謹向過去一年辛苦投入編纂《和AI做朋友》的所有教師與助理團隊致敬及致謝，相信我們的努力終將能提升臺灣回應AI時代挑戰的競爭力。

教育部部長

潘文忠

謹誌

中華民國108年8月

編者的話

為推廣人工智慧教育並於中小學奠基，培育學生成為未來引領AI世代的科技人才，教育部籌組學者專家團隊，執行「人工智慧技術應用與人材培育計畫—中小學分項：中小學推廣教育計畫」（以下稱為本計畫），發展中小學人工智慧教材教案示範例。

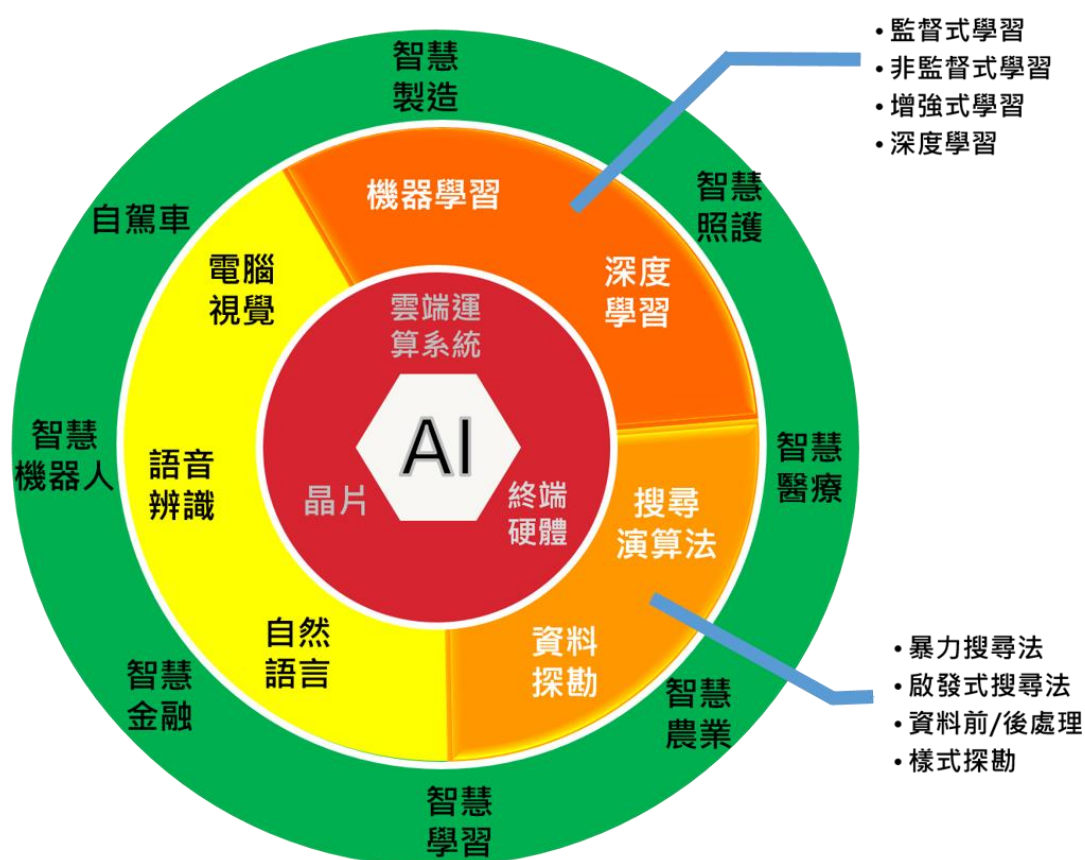
編輯團隊

為中小學設計適合的人工智慧教材，需要考量各年段學生的數理知識程度以及所要傳遞的人工智慧內涵，以適當規劃教材所要涵蓋內容的深度及廣度。計畫辦公室邀請相關領域的大學教授提供人工智慧的知識性內涵，並邀請教育現場第一線的教師，將上述人工智慧的內涵轉化為適合中小學學生及能配合十二年國民教育基本課程綱要的教材及教案。教材教案編輯團隊成員包括來自國立臺南大學、中華大學、國立成功大學、國立臺灣科技大學和南臺科技大學等大學教授，以及桃園市笨港國小、國立臺灣師範大學附屬高級中學國中部、國立臺南第二高級中學等現場教師共同合作，產出國小、國中、高中三階段的人工智慧教材教案。

為什麼是示範例？

人工智慧領域涵蓋的範疇甚廣，依據十二年國教課程綱要，資訊科技科目列為國、高中必修科目，其授課時數在國中階段每週1節、高中階段為2學分，高中可以選修方式列入課程，國中可採融入課程方式，國小則可於彈性課程實施或採融入方式進行教學，但以如此有限的時數要教授如此廣的內容確有其難度。團隊經過不斷思索研議後，決定先挑選具代表性的機器學習主題發展教材教案，若第一線教師有意願將此教材引入教學中，並在使用後覺得設計理念符合教學所需，願意投入後續發展，本計畫將逐步擴大教材發展團隊。這是「示範例」的第一個概念。

這波全球人工智慧熱潮大爆發起因於，Google的人工智慧圍棋軟體AlphaGo打敗世界棋王的新聞事件，進而讓AlphaGo背後的「深度學習」技術為世人所關注，並成為人工智慧的代名詞。事實上深度學習是機器學習的一個分支，機器學習歷經多年發展，體系相對完整，因此本計畫先以機器學習及深度學習兩主題發展教材教案，並預計在下一階段中，發展搜尋演算法及資料探勘等主題。這是「示範例」的第二個概念。



我們選擇的主題

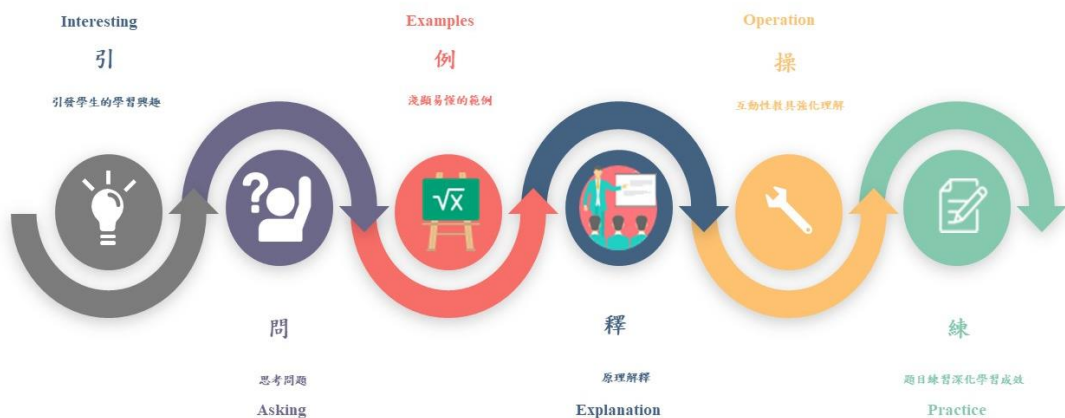
因人工智慧範疇廣袤且分支眾多，團隊依據運算資源及硬體設備、核心技術面、應用面等三個面向大致劃分出人工智慧現有範圍，以支持人工智慧運作最基本需求的運算資源及硬體設備為基底，向外延伸出機器學習、深度學習……等核

心技術面，而最外圍則是依據上述基礎所發展至生活中的各項應用。

本團隊考量高中學生已經具有足夠的數理基礎，所以著重介紹四種機器學習類型，包含：監督式學習、非監督式學習、增強式學習與深度學習。至於國中和國小教材，考量到學生的數理程度，著重於引起學生學習興趣並介紹人工智慧於生活中的應用，因此，除了簡介人工智慧及提供背景知識之外，主要課程內容是機器學習領域中最具代表性的監督式學習及非監督式學習介紹。

編輯理念說明

本計畫利用引、問、例、釋、操、練的順序設計教材，先導入多媒體來「引」發學生的學習興趣，再以提「問」的方式引導學生思考，並透過範「例」的講解，讓學生歸納重點後進行原理解「釋」，藉由互動性教具的「操」作強化理解，最後提供「練」習題目以深化學生的學習成效。



除了發展核心知識教材以外，我們也發展實作教材，提供高中學生開發人工智慧應用程式的機會，讓高中學生能透過實作對人工智慧有更深一層的體驗。實作教材使用 Jupyter Notebook 為執行環境，可以直接在 Google Colab 平台操作；學習完成實作教材的學生，若想藉由製作更大型人工智慧專案，以求更深入鑽研人工智慧技術，可透過指導老師申請使用教育部建置的 AI 虛擬雲端平台。另外，為提升國中及國小學生的學習興趣與成效，國中及國小階段教材有搭配實體教具，

讓學生能以動手操作的方式體驗機器學習的運作流程。此外，本教材也提供目前在中小學使用率較高的程式設計語言Scratch範例，讓學生進行操作演練。

給教師的話

對人工智慧感興趣的教師夥伴們，想將人工智慧融入教學中卻苦於始終找不到著力點嗎？本團隊除了開發高中、國中、國小各階段的教材之外，也依據教材開發出相應的教案，提供給對人工智慧教育感興趣的教師使用。所有階段的教材教案皆經過本團隊於現場實施教學後，檢討回饋再加以修正。依據團隊實際實施教學經驗，若是教師想要於現場進行教學，提供以下建議：高中階段教材所編撰的內容較深入，且教材的內容單元有搭配相應的上機實作教材，可採列入多元選修課程實施教學；國中階段教材可於彈性學習課中實施教學，並建議於國一或國二開設，若是上課時數不足，可以挑選數個活動任務，融入各領域課程之中，以符合跨領域（科技領域）精神，也可以開設一個學分的社團活動課程，挑選其中18小時進行一學期的AI人工智慧體驗課程進行教學；國小則可以採跨領域融入的方式進行教學，或是於彈性課程中實施，若是上課時數不足，則可以挑選適合的章節授課，以符合實際需求。

給學生及家長的話

對人工智慧有興趣的家長及同學，歡迎你們即將共同參與台灣教育史上嶄新的一頁。團隊所開發的教材包括：高中、國中、國小三個階段，考量現實情況，各階段學生大多是第一次接觸人工智慧，所以這三個階段的教材不具連續性，只有內容深淺的差異，可以獨立使用，且皆是由人工智慧簡介開始，由淺入深介紹人工智慧的核心知識及應用層面，讓各階段從未學習過相關知識的學生，能對人工智慧有基本認識。從未接觸過人工智慧領域的學生，若是有興趣想要學習此教材內容，有以下建議提供參考：若是高中學生，建議學習過高中一年級的數學課程，因高中階段教材需要用到向量、變異數等數學概念，若是有程式設計基礎者，

則在完成基本實作課程後，可至教育部AI雲端平台進行專案製作；若是國中學生，則建議國一以上，具備判別資料特徵的能力，以及最短距離的概念；若是國小學生，則以高年級較為合適，建議具備數數及簡單統計的概念，後面章節需要利用到表格判讀數據，及直尺作圖的能力。

希望這份教材可以成為同學認識人工智慧的啟蒙，並期待每位學習過此份教材的同學，能成為未來AI世代的人才。

結語

希望此份教材能對想將人工智慧概念帶入教學的老師有所幫助，並對將人工智慧教育往下扎根於中小學有所貢獻。由於教材設計牽涉的層面廣袤且相當繁複，難保不會有謬誤之處，若各位先進及對人工智慧教育具熱忱之教師及學子，發現有任何謬誤疏漏之處，祈請不吝指正，讓本教材能日趨完備。最後要感謝在此教材教案不同發展階段的審查委員們，由於他們所提供的寶貴專業意見，讓此教材教案得以更臻完備。

李 建 樹

國立臺南大學資訊工程學系教授

中華民國108年8月

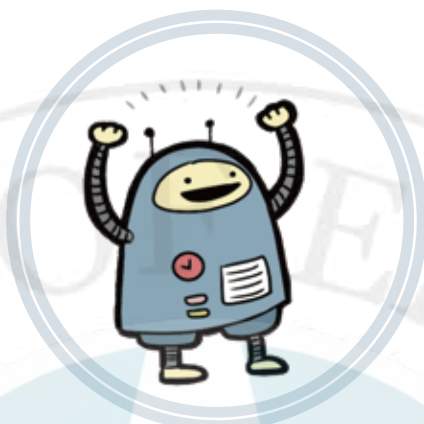
第一章：	解開AI身世之謎－人工智慧簡介	2
人物介紹		2
教材特色		2
1-1	人工智慧與生活	3
1-2	人工智慧的發展史	12
1-3	人工智慧與機器學習	19
第二章：	從不同中發現相同－人工智慧的背景知識	21
2-1	惱人的數據	21
2-2	揪出同夥人	28
2-3	芝麻開門－快來探聽通關密碼	31
第三章：	從樣本中大膽預言－監督式學習	36
3-1	「向左走向右走」－決策樹(Decision tree)	36
3-2	「誰上線?誰離線?」－線性回歸	48
3-3	「線性?或非線性?」－感知器(Perceptron)	56
3-4	「看誰勢力大」－KNN分類器	62
第四章：	從混沌中搜尋同類－非監督式學習	65
4-1	「是我族或他類?」－K平均值分群法	65

人物介紹



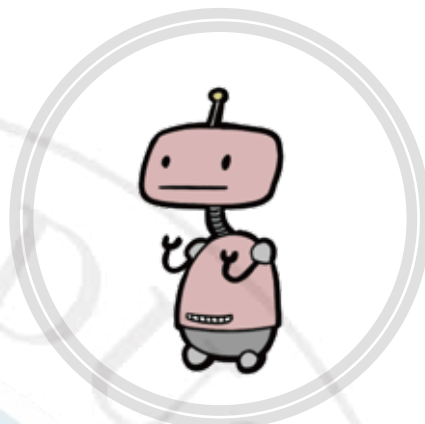
小叮

家有機器人的小主人



阿圖

憨厚忠誠的機器人

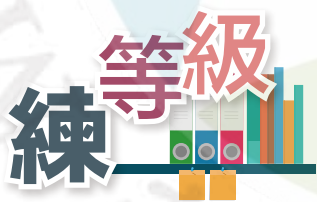


靈兒

勤快機靈的機器人

教材特色

「透過任務學AI」是本教材的特色，為拉近國中生與AI的距離，本書各單元均設計專屬任務，搭配個人獨自探索的「練等級」、小組破解任務必備的「長智慧」、收集AI智慧的「撿裝備」和獲取獨門秘訣的「神救援」等，大家趕快隨著阿圖一起出任務，練成無敵AI達人吧！！



個人進行腦力激盪與嘗試



大腦可以靈光乍現的密碼

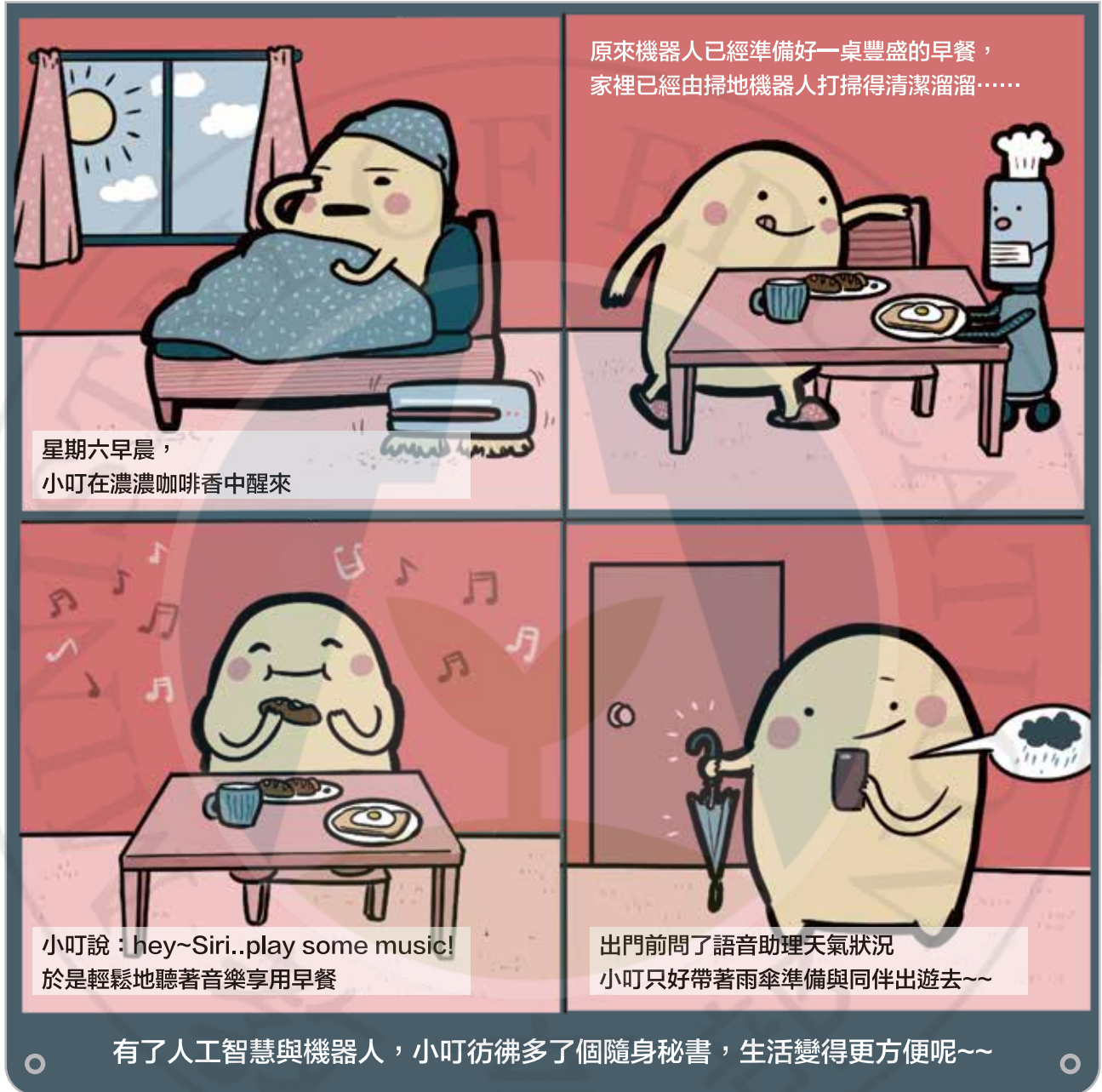


小組可以分工搶奪的秘辛



提供困難問題暗示與解密

1-1 人工智慧與生活



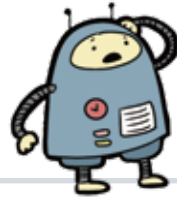
等級 練

想想看，在我們一天生活當中，還有哪些工作需要使用電腦人工智慧呢？

小叮家有 ➡ _____、_____、_____。

我想要有 ➡ _____、_____、_____。

第一章：解開AI身世之謎－人工智慧簡介



阿圖出任務



關卡一 ● 尋找機器人

請同學四人一組，分享彼此想到的機器人並完成下表。

組別：第	組員	機器人	工作	厲害招式
	例	廚師機器人	麵包、煎蛋餅	顏色辨識、溫度監控
	1			
	2			
	3			
組	4			



關卡二 ● 機器人創意故事

請各小組發揮創意，用這四個機器人的特質，編成一個100字左右的小故事，別忘了分頭去『長智慧』和『撿裝備』喔！

Blank writing area for students to create a story.



甚麼是人工智慧

人工智慧的英文是「Artificial Intelligence」，因此簡稱為「AI」，意思是讓機器具有人類的智慧。人工智慧就像3C產品一樣，已經普遍存在我們的生活當中，也慢慢地影響我們的生活，以下將介紹人工智慧在智慧視覺、智慧聽覺、智慧醫療等方面的應用。

檢裝備

【裝備一】智慧視覺

智慧視覺：包含「人臉辨識」、「植物辨識」、「動物辨識」、「車牌辨識」以及「車輛自動駕駛系統」等。

1 人臉辨識：

人臉辨識是在影像辨識中，使用最廣泛的一門技術，也因為技術越來越純熟，人工智慧程式在臉部辨識的準確度上，幾乎已經超越人類所能。舉凡手機上的Face ID臉部解鎖、拍照臉部對焦、手機相簿人物分類、智慧門鎖、課堂點名(如圖1-1、圖1-2)等，皆廣泛使用到人臉辨識技術。

人臉辨識的原理，是將臉部的圖像輸入至系統，再與原本擁有人臉的資料庫做比對，即可識別出身份。有越來越多人家裡使用智慧門鎖，出門不需帶著大把鑰匙；公司行號使用人臉辨識作為進出公司門禁管制、行動支付使用臉部辨識增加其交易安全性。關於臉部辨識有一個知名的案例：日本知名歌手宇多田光為避免演唱會門票遭黃牛介入，因此歌迷買票及入場都必須使用人臉辨識系統，以確保本人使用。



圖1-1 上課簽到簽退之實例

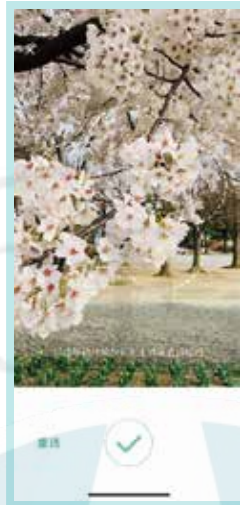


圖1-2 使用人臉辨識系統點名

第一章： 解開AI身世之謎 – 人工智慧簡介

2 植物辨識：

依據植物的生物特徵，例如形狀、顏色、數量、排列方式等，將欲判別的照片匯入，系統即可自動識別出植物的種類及名稱。



選取手機中的花卉照片，框框對準要辨識的花卉，系統會與資料庫中的照片作比對，自動辨識出最符合的花卉名稱。



圖1-3 使用APP自動辨識花卉種類

3 動物辨識：

在茫茫照片海中，我們可以將相片中的動物種類做分類，並能做出正確的檢索。下圖左顯示相簿能自動標示出拍攝時間、地點，下圖右表示從相簿中能搜尋出有關所有「貓」的照片。

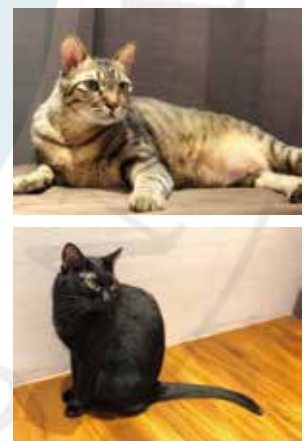


圖1-4 在手機相簿輸入關鍵字“貓”，即可列出所有相關的照片

4 車牌辨識：

以前車子開進收費停車場時，需要所取票卡或代幣，離開前拿著票卡或代幣到繳費機繳費，票卡若不小心遺失還得賠錢。現在有了車牌辨識系統，車輛進出停車場，改用車牌辨識計算停車時間及費用，用來取代人工票券或代幣，再也不用擔心票卡遺失的問題了，而且進出停車場更有效率。圖1-5是車牌辨識停車場的運作示意圖：



圖1-5 停車場之車牌辨識系統使用流程示意圖

5 車輛自動駕駛系統：

車輛自動駕駛系統是當行車時，由電腦系統自動偵測周遭車輛、自動停車、自動根據車況控制車速，可減輕人類開車的疲憊感。最著名的自動駕駛車－特斯拉Tesla，它不但可配合交通狀況調整車速、並保持在車道內行駛、並有可自動變換車道（不須駕駛者控制）、自動停車等功能。雖然自動駕駛帶給駕駛者許多方便之處，但駕駛人仍須在行駛時提高警覺，隨時注意路況，才可確保安全。

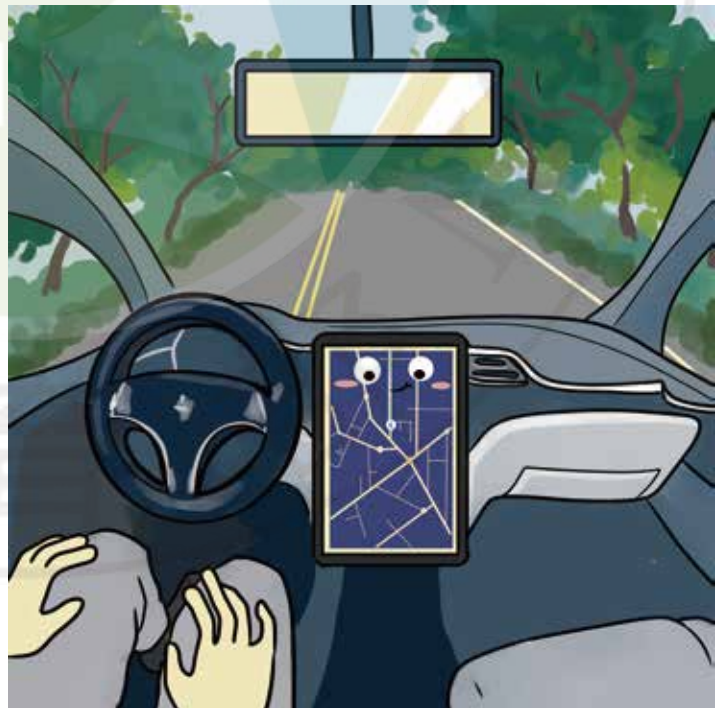


圖1-6 自動駕駛示意圖



【裝備二】智慧聽覺

智慧聽覺：包含語音辨識、語音助理、歌曲辨識等。

語音辨識可將聲音轉換為文字，針對不擅長打字的人，真是一大福音！而語音助理我們可視它為一個虛擬的小幫手。

常用的語音助理，例如：Apple 手機中的Siri、Google手機中的Google Assistant、Amazon的智慧音箱Echo、小米的AI音箱、Apple智慧音箱HomePod等，除了提供語音撥電話、語音問答等功能之外，也可請語音助理設定鬧鐘、查詢天氣、喚醒起床、甚至餐廳訂位等，儼然成為一個隨身秘書了呢！



圖1-7 使用Siri詢問明天天氣及設定鬧鐘

在咖啡館聽到一首好聽的歌，想回家好好聆聽～這時只要打開手機App（例如Shazam、SoundHound等），它馬上可以告訴你這首歌的資訊，是不是很方便呢？



圖1-8 使用Shazam辨識歌曲APP

智慧醫療：包含遠端醫療、快速診斷、機械手術、老人照護等。

人工智慧對於醫療發展有很大的助益，從早期血液檢查、細胞檢查、斷層掃描檢查等以往得知結果需耗費數小時或數天，有了人工智慧的幫助，可以縮短疾病判斷的時間，掌握治療的黃金時間。

此外，在醫療手術、健康管理、遠端醫療或疾病預防等方面，也因人工智慧蓬勃發展而使醫療更周全、更有品質與效率。隨著人口老化情形日益嚴重，因應高齡化社會來臨，人工智慧在家庭或是老人照護上的發展也有了創新。例如：近年日本RIKEN研究機構研發出一款照護型機器人-Robear，其臉部看似一隻可愛的熊。雖然是個冷冰冰的機器人，卻可以給予需要被照護者陪伴以及提供一般的照護協助。

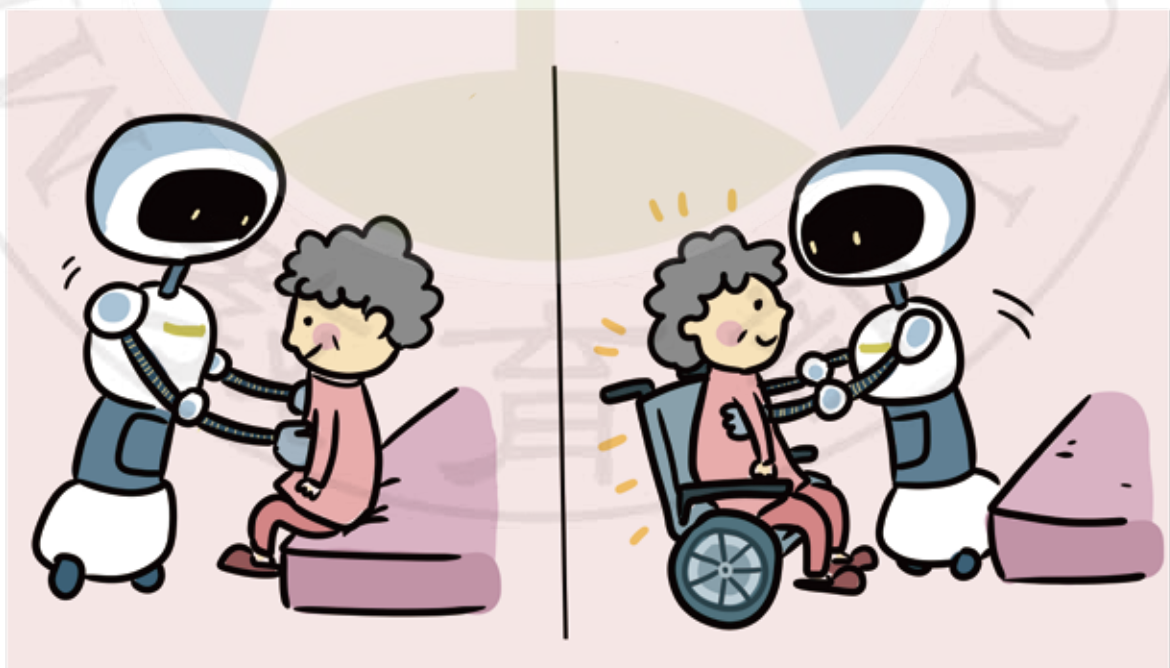
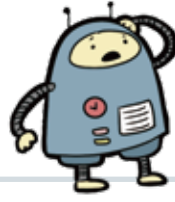


圖1-9 照護型機器人之功能示意圖（協助患者上下床或輪椅）



阿圖出任務



關卡三 • 小叮賞花記

下面這篇文章是小叮賞花遊記，看完後，有一項配對任務等著你喔！

¹ 昨天爸媽提議星期六下午全家一起搭乘新買的無人駕駛車到國家公園賞花。² 我趕緊拿起手機詢問星期六下午國家公園是否會下雨，知道「天氣晴朗」我就使用手機語音助理開始規劃行程。

到了星期六下午全家一起搭車上山。³ 進入停車場前，柵欄暫停一下掃描車牌才打開讓車子進入停車場。我們在售票機前買票，居然不用取票耶！只要結帳完對著售票機拍照⁴ 直到入口處看了一下鏡頭，門就自動打開了，一家人興高采烈賞花去。

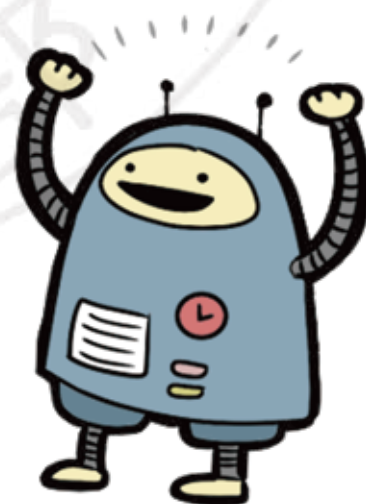
國家公園開滿了各式各樣爭奇鬥艷的花，我拿起手機拍照。這天好多人也帶著家裡的寵物一起來賞花看到不知名的花，⁵ 爸爸說可以打開App查詢是甚麼花這麼美，我透過App終於認識國家公園各種美麗的花，我還拍了好多寵物的照片。

⁶ 回家整理照片時用了手機相簿的檢索功能，搜尋出在公園拍的狗兒們並與愛狗的媽媽一起欣賞，很开心全家的賞花之旅留下美好回憶。



請將下列1-6情境，找出最適當的人工智慧技術，並用直線連起來

情境	1	2	3	4	5	6
人工智慧技術	<input type="radio"/> 植物辨識	<input type="radio"/> 臉部辨識	<input type="radio"/> 自動駕駛系統	<input type="radio"/> 動物辨識	<input type="radio"/> 語音助理	<input type="radio"/> 車牌辨識



1-2 人工智慧的發展史

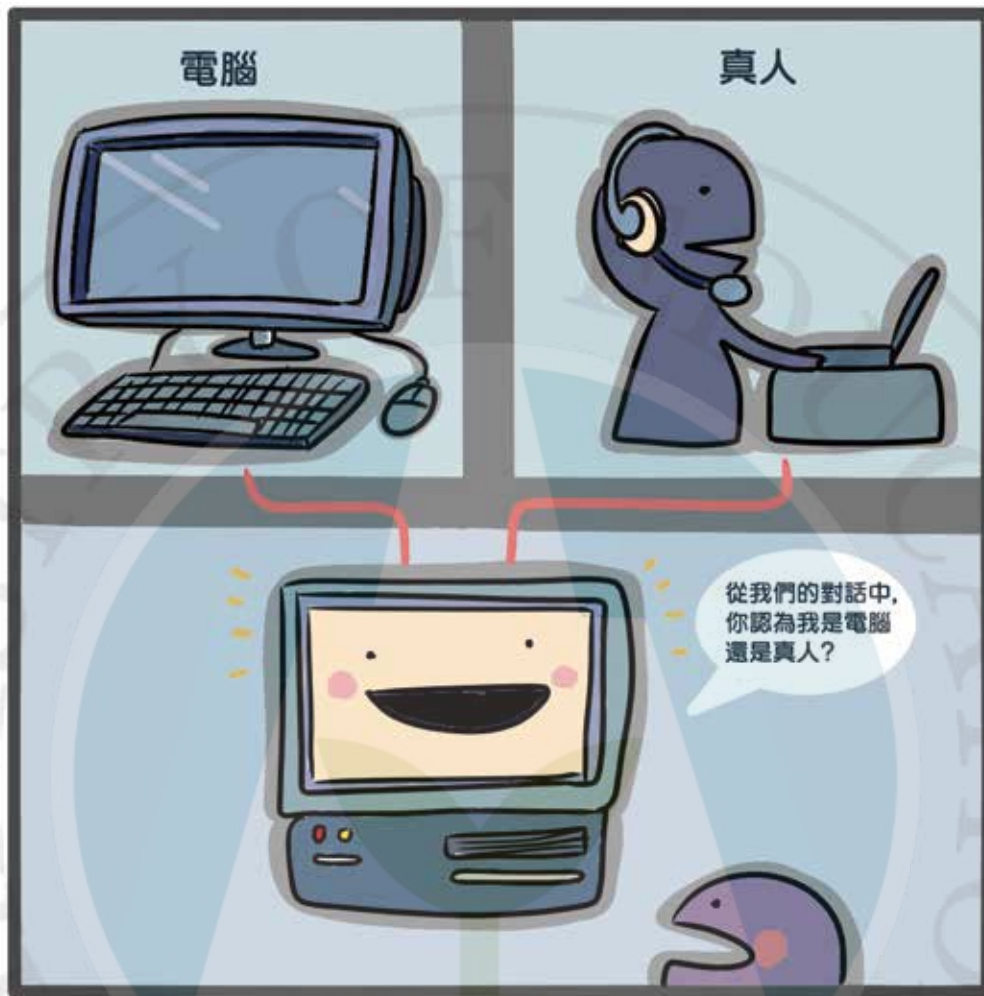


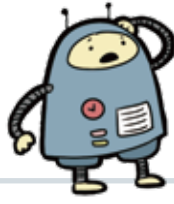
圖1-10 圖靈測試示意圖

你聽過「圖靈測試」嗎？這個實驗是由英國電腦科學家艾倫圖靈所提出的，目的是想知道機器是否具有人工智慧。



如果你有一通未接來電，並且懷疑是小叮家的秘書機器人打給你的，在回電話以前先想想看，你要問什麼問題才能測試出接電話的是小叮本人？還是秘書機器人呢？

我想要問 → _____



阿圖出任務



關卡四 • AI達人PK賽

相信現在的你已經增長了許多關於人工智慧的知識並蒐集許多裝備了吧！？
以下有關人工智慧的問與答，讓我們來驗收一下吧！

- 什麼是人工智慧？
- 圖靈測試是如何運作的呢？
- 「專家系統」主要功能是什麼？
- 什麼是「弱人工智慧」？
- 舉例說明人腦和電腦的競賽。

第一章：解開AI身世之謎－人工智慧簡介



關卡五 • 人工智慧大未來

儘管人工智慧為我們的生活帶來許多便利，但是也對人類的生活造成某些部分的影響。人工智慧專家李開復曾在「人工智慧來了」一書之簡介中提及：

『人工智慧時代，將有50%的工作逐漸被取代！？』

小叮看到這段話開始眉頭深鎖，¹ 你認為有哪些工作是可能會被取代呢？² 哪些工作是不容易被取代的呢？³ 小叮正煩惱著未來該學習甚麼才不會被淘汰？⁴ 應該具備甚麼特質才能在AI社會中立足呢？

請各組一起集思廣益給小叮一點建議，並完成下表！

組別：	討論題
第	1. 可能被取代的工作：
組	2. 不容易被取代的工作：
	3. 未來可努力的方向：
	4. 應該要具備的特質：

艾倫圖靈（Alan Turing，1912-1954）（以下簡稱圖靈）是出生於英國的電腦科學家，同時也是數學家、密碼學家、生物學專家，以及世界級馬拉松選手。在第二次世界大戰期間，圖靈在英國政府的情報機構負責密碼破解的工作。其中最著名的事蹟是領導團隊設計出一套破解德軍密碼的系統－「Enigma」密碼機，這個系統設計讓當時英國對德國的軍事行動得以瞭若指掌，並使該次戰爭至少提早兩年結束，且至少讓1400萬人倖免於難。



艾倫·圖靈
1912-1954年

由於圖靈所發表的「圖靈測試」與「圖靈機」為人工智慧領域奠定了基礎，圖靈因此被譽為「人工智慧之父」。

檢裝備

自從「人工智慧」這個名詞誕生之後，人們對它便抱以無限的想像與期待。其實人工智慧的發展並非一直都很順利，從發展之初到現在60多年之間，大約經歷兩次低潮，它的發展興衰示意圖如右圖所示。

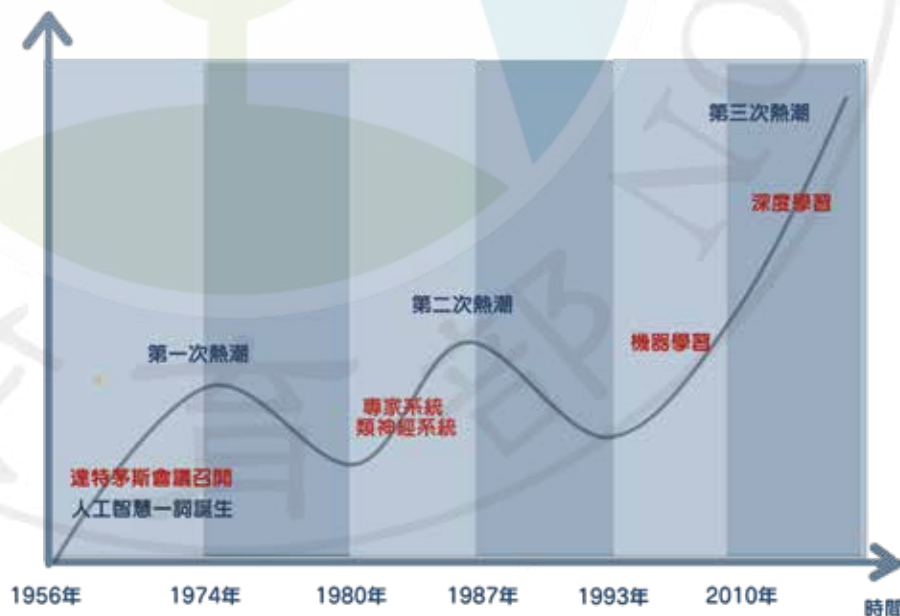


圖1-11 人工智慧的興衰示意圖

第一章： 解開AI身世之謎－人工智慧簡介

第一次熱潮（1956-1974）：

「人工智慧」一詞第一次出現是在1956年於美國召開的達特茅斯（Dartmouth）夏季人工智慧研究計畫會議上。在這次會議之後的數年之間，人工智慧大躍進。當時的電腦科學家們對於人工智慧擁有很高的期待，也充滿研究熱忱。

1963年，美國國防高等研究計畫署（DARPA）提供大量資金給麻省理工學院（MIT），也因此培養出一批人工智慧專家，對人工智慧初期的發展有深遠的影響。在這個計畫中最著名的莫過於1964年Joseph Weizenbaum教授在麻省理工學院設計出來的第一個聊天程式ELIZA。



圖1-12 對人工智慧充滿熱忱的科學家

ELIZA是模擬心理醫生與人對話，讓使用者誤以為在和人交談，而不是電腦程式。ELIZA並沒有建立完整的句子結構，也沒有機器學習，它是從人類的問題中找出預先載入的關鍵字，運用自己現有的對話知識，若找到符合者，就回應相對設定好的句子；反之則從預設的通用答案中選擇一個回覆。ELIZA只能處理簡單的對話，它雖然不像現今的語音助理般聰明，但卻是現今自然語言教科書中最經典的例子之一。



圖1-13 聊天程式ELIZA示意圖

從人工智慧興起之初，電腦科學家們樂觀地認為：具有完全智能的機器將會在20年內出現，並期望能夠藉由電腦解決問題，但礙於電腦的發展速度及運算能力，遠不及人們的需求，導致人工智慧的研究遇到瓶頸而停滯不前。因此在1970年代人工智慧的發展遇到第一次低潮。

第二次熱潮（1980-1987）：

隨著電腦硬體的發展與普及，專家系統與類神經網路的研究在80年代興起，人工智慧的研究逐漸推廣到實用層面。這段期間誕生第一套「專家系統（Expert system）」－DENDRAL，該系統是由專家系統之父－Edward Feigenbaum團隊所發展出來，主要目的是使用化學知識來鑑別未知的化學有機分子。

專家系統主要是由大量的知識庫及推理規則構成，模擬專家解決複雜問題並進行決策。例如生病時，醫生是根據病狀判斷可能的疾病，並採取對應的治療方式。使用專家系統診斷病情時，專家系統會依照病狀參考資料庫，並提示可能的疾病。如：咳嗽、流鼻水、頭痛等三個症狀，專家系統可能會判斷為「感冒」，並且提出感冒的治療對策。

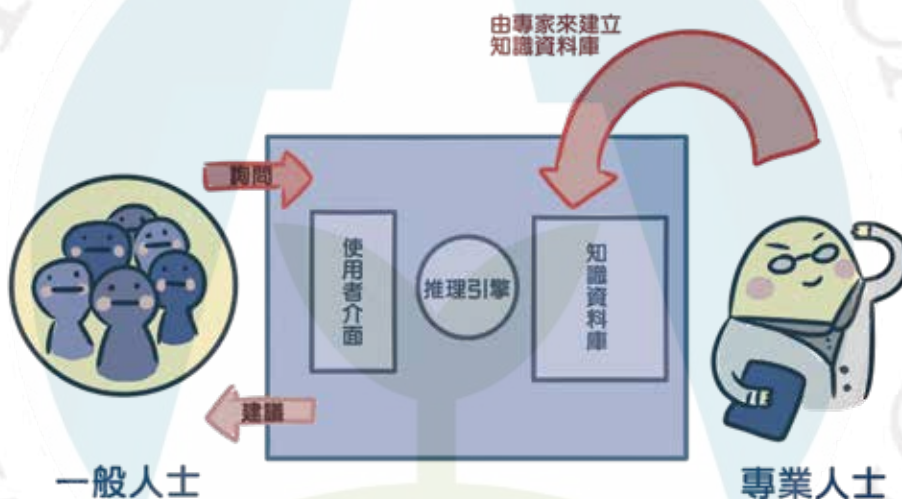


圖1-14 專家系統示意圖

專家系統其實是眾多「若A則B」的規則，若輸入的規則與知識越多，人工智慧就會變得越聰明。當時的專家認為龐大的知識庫需要不斷堆砌與大量的維護，因此投入大量資金。

類神經網路在這段期間也有新的進展，但1980年代末期因專家系統及類神經網路開發及維護費用昂貴，應用範圍不夠廣泛，因此投入的資金驟減，導致發展遇到瓶頸，人工智慧進入第二個低潮。

第一章： 解開AI身世之謎－人工智慧簡介

第三次熱潮（西元1993年-至今）：

在第二次人工智慧低潮之後，大約1990年代，人工智慧的發展僅侷限於弱人工智慧。所謂弱人工智慧是只能模擬人類的思維，但不具有完整的人類感官及認知能力，目前多數標榜人工智慧產品都屬於弱人工智慧。

隨著網路崛起，電腦硬體也有很大的進步，因此電腦的運算能力及處理資料的能力大大提升。近幾年使用大量數據（Big Data）來分析資料並進行機器學習的方法越來越多，也使得人工智慧在分析或解決專門問題上有長足的進步。有許多公開競賽使得人工智慧的發展更具競爭力也更被世人看見。例如：IBM的Waston益智問答競賽。

2016年AlphaGo電腦圍棋以4：1打敗曾獲得世界職業圍棋公開賽冠軍多次的韓國棋王。這些知名的人腦與電腦競賽，讓過去人們認為很難突破的問題，藉由硬體的進步與電腦運算能力的提升，讓人們見識到深度學習在人工智慧領域上所呈現的成果。

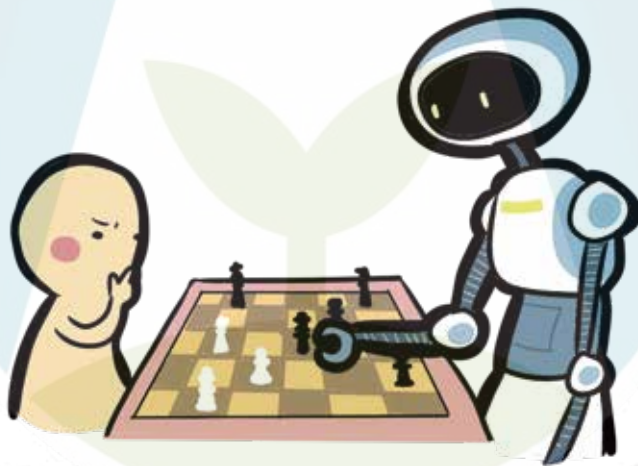


圖1-15 小叮與機器人比賽下棋

第三次的人工智慧熱潮來自深度學習的發展與成效，不管在商業、工業、農林漁牧或教育等各領域都有相當令人振奮的效果，甚至超越人類，也因此這幾年深度學習仍然在人工智慧領域活躍著，人工智慧的第三次熱潮也正持續發展中。

1-3 人工智慧與機器學習

完成前面的五個關卡之後，是不是對人工智慧如何運作產生興趣呢？是不是想要進一步了解人工智慧的奧秘呢？先不急，後面章節將會逐步帶領大家認識人工智慧的內涵，進入人工智慧的異想世界。進入第二章之前，讓我們先來認識人工智慧的相關名詞及基本知識吧！

「深度學習」是大家最耳熟能詳的人工智慧專有名詞之一，一般人往往將深度學習跟人工智慧畫上等號，但事實上深度學習只是機器學習的一個分支，而機器學習又是人工智慧領域的一個分支，三者的關係如圖1-16所示。知道他們之間的關係後，日後就比較不會被這些名詞給弄糊塗了。

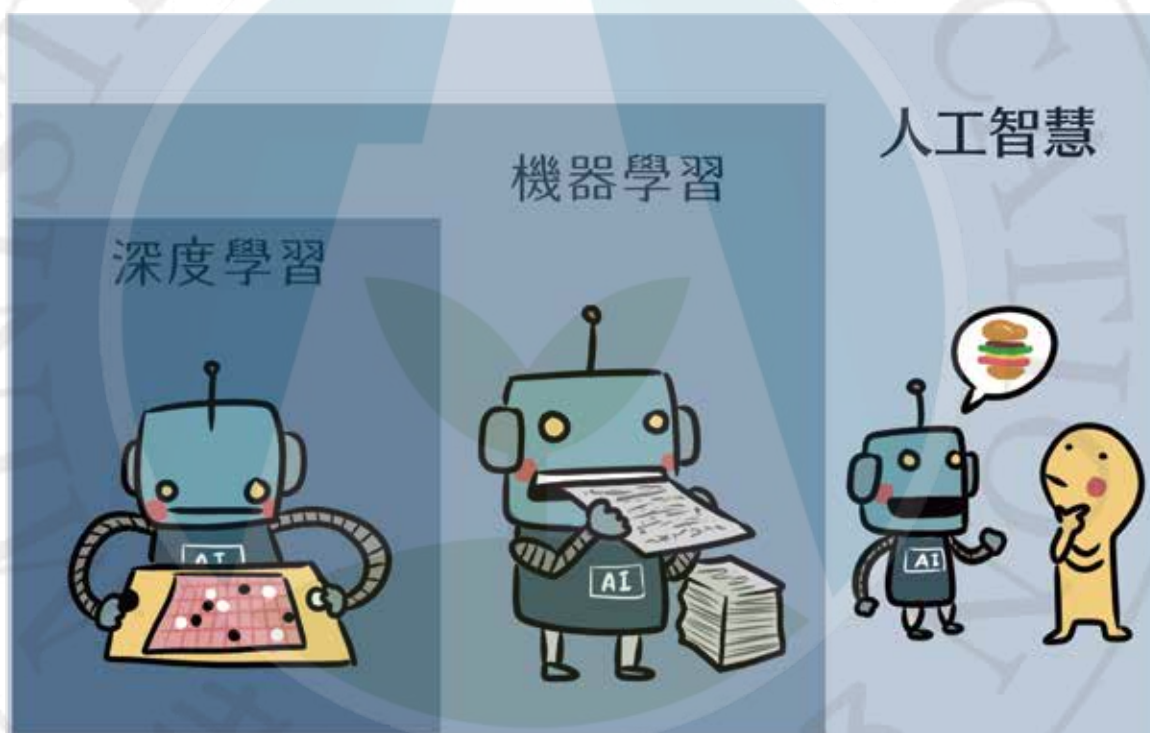


圖1-16 深度學習、機器學習與人工智慧關係圖

由於人工智慧領域所涵蓋的範圍很廣，要想完整涵蓋整個人工智慧並不容易。最近的人工智慧熱潮大爆發，主要的原因之一是來自深度學習令人驚艷的表現，而其隸屬的機器學習歷經多年發展，體系相對完整，因此本教材會著重在機器學習領域中最具代表性的二種機器學習類型介紹，即監督式學習和非監督式學習。

接下來的第二章將針對機器學習的眾多技術中，較常使用及較基本的知識做介紹，第三章到第四章則是針對上述的二種機器學習類型做進一步介紹。以下是第二章到第四章的內容摘要性說明。

■ 第二章：從不同中發現相同 – 人工智慧的背景知識

這一章是讓同學們先了解人工智慧的基本背景知識，例如：資料要如何收集及整理？資料特徵是什麼？以及一些名詞的介紹。

有了這些背景知識之後，便可以進一步學習人工智慧的兩種機器學習類型。

■ 第三章：從樣本中大膽預言 – 監督式學習

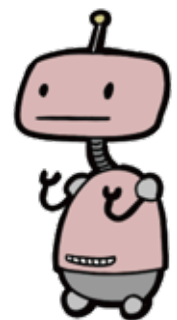
為方便理解，我們可以把監督式學習視為「有老師指導」的學習方法，因為老師擁有正確答案(或稱標記)，所以在訓練階段可以提供每個訓練樣本所對應的標記資訊，根據訓練樣本的對應標記資訊所設計的機器學習方法就叫做監督式學習。

在這一章，同學們將會學到監督式學習中的決策樹、線性回歸及感知器等方法的運作原理。

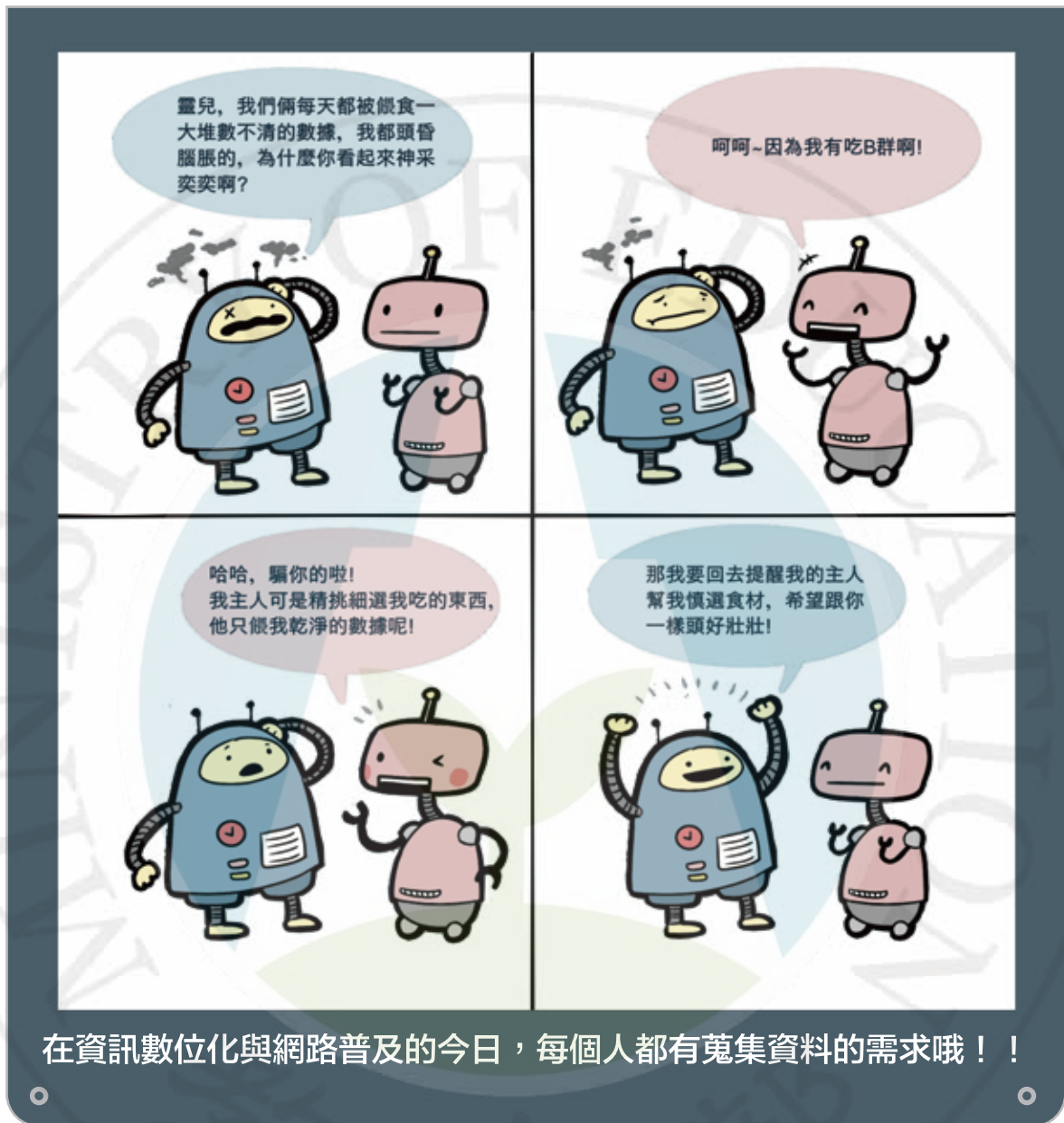
■ 第四章：從混沌中搜尋同類 – 非監督式學習

在現實運用上，採用監督式學習便要取得每個訓練樣本所對應的標記資訊，這部份往往需要付出可觀的人力成本。以訓練機器辨識動物影像為例，要對每張訓練用的動物影像標記影像內容是哪種動物，便需要請專人做這件事。為避免此一限制，便有非監督式學習方法被提出，相對於監督式學習有老師指導，非監督式學習就是一種「沒有老師指導」的學習方法，也就是說，在訓練階段無需提供訓練樣本所對應的標記資訊，也由於欠缺此這些資訊，因此通常會就訓練樣本彼此之間的相似性進行聚類分群。

在這一章，同學們將會學到K-平均演算法以及了解如何讓電腦自己去分群。



2-1 惱人的數據



等級 練

小叮最近發現把電子發票條碼上傳到某網站，竟然可以獲得遊戲點數！想想看，發票上面有哪些數據呢？這些數據可以怎麼整理呢？

發票上有 ➡ _____

整理方法 ➡ _____



阿圖出任務



關卡一 ● 尋找金牌選手

如果學校想推派四位選手參加2020年全國中等學校國中組男子100公尺、男子5000公尺、女子100公尺、女子3000公尺田徑比賽，但是全校有1000多人，目前體育老師提供了每位同學每學期的短跑100公尺和迷你馬拉松3500公尺的成績數據，試想想，要如何利用這些數據找出代表選手呢？

請將討論的結果寫在下表：

	組員
組別：第 組	

這一章是讓同學們先了解一些人工智慧的基本背景知識，例如：資料要如何蒐集及整理？資料的特徵是什麼？以及一些專有名詞的介紹。等大家了解這些知識之後，接下來才可以進一步去學習人工智慧的運作原理喔！

檢裝備

數據大集合

在人工智慧領域裡，想要讓電腦具有處理龐大數據的能力，首先就要準備大量的資料。到底要去那裡尋找資料呢？以下是幾種常見的蒐集數據方式：

【裝備一】

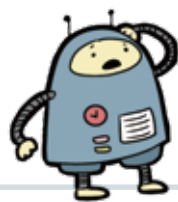
1 政府公開資訊：

政府單位將資訊公開在網路上，使用者只要遵守網站資料開放相關規定，針對個人需求，便可以在網路上擷取自己所需資料。政府資料開放平臺網站中，包含求學進修、求職就業、休閒旅遊、生活安全及品質等各種琳瑯滿目的資訊，皆開放自行下載。例如：細懸浮微粒（PM2.5）、各縣市人口統計等在網站中都可以找到。右圖為政府資料開放平臺網站之首頁：

<https://data.gov.tw/>



圖2-1 各行各業的人都可透過政府資料開放平臺下載自己所需資料



阿圖出任務



關卡二 ● 數據會說話

請同學使用瀏覽器連結到政府資料開放平臺網站（網址：<https://data.gov.tw/>），然後各組選定一個主題，點進去看看是否有合適的資料可分析，¹你會怎麼分析那些數據？²或者想從分析後的數據做甚麼應用？請各組討論後與同學分享。

組別：第	組員
組	<p>1. 請寫下資料分類及主題（例如：縣市觀光景點之參觀人次）</p>
	<p>2. 該數據經過分析後，可以拿來做甚麼應用？</p>



【裝備二】

2 網路爬蟲 (Web crawler) :

許多人會使用網路搜尋引擎（例如Google Chrome、Firefox、Microsoft Edge等）尋找資料，可能是為了做報告、可能是為了找交通路線、可能是為了查電影上映時間，也可能是為了找美食、找旅遊景點等。

當資料較少的時候，可以一筆一筆尋找並分析整理，但是當資料多到上百筆甚至上千筆時，是不是有更好的方法可以節省時間呢？有的，這時可以使用「網路爬蟲」程式。「網路爬蟲」又稱為「網路蜘蛛」，它就像蜘蛛網一樣，由中心主題往外延伸。網路爬蟲的原理是透過網頁超連結尋找網頁，從初始網站某一頁面開始讀取網頁內容，找到關鍵字的超連結網址，再透過這些超連結尋找下一個網頁，如此循環，直到瀏覽所有網頁，並將所需資料擷取，這種技術就叫做「網路爬蟲」。



圖2-2 網路爬蟲 (網路蜘蛛)



【裝備三】

3 公司行號API :

API是Application Programming Interface（應用程式介面）的縮寫，藉由資料交換，可使企業與合作廠商之間資料流通更方便。

部分公司為了讓軟體開發業者可以為公司額外開發更有用的應用程式，藉由開放的API，讓開發者可以利用公司提供的資料，進行各種應用程式開發。開發者必須遵守業者的API開放規章。例如：知名的網路搜尋引擎公司Google、社群網站FaceBook、Twitter、網路相簿Flickr或影音平台Youtube等，這些平臺業者都有提供自己本身的API讓使用者可以下載平臺資料來開發更多的應用。



圖2-3 社群網站提供API供使用者下載



【裝備四】

4 其他 :

除了上述資料蒐集的方法之外，還有許多其他的資料來源可蒐集資料。例如：使用網路攝影機蒐集影像資訊進行人臉辨識、物品分類、場景辨識；使用錄音設備蒐集聲音檔進行語音辨識、歌曲辨識，蒐集資料的方式愈來愈多元，也使得人工智慧可以應用的領域更加廣泛。



數據大整理

數據整理：

當我們蒐集到各種資料之後，並不代表全部資料都可使用。利用各種方法蒐集到的資料，必須經過整理、分析才能成為有效的資訊，也才能產生它的價值，因此錯誤或無法使用的資料，必須在進行機器學習之前先進行處理。

在電腦的領域中，有句話叫做「GIGO」，意思是「垃圾進、垃圾出（Garbage in, garbage out!）」。如果我們輸入電腦的資料是錯誤的、或沒有意義的，那無論經過甚麼形式的資料處理，電腦回報給我們的資料也會是錯誤的、沒有意義的訊息，這也顯示輸入正確資料的重要性。

蒐集好的資料在還未整理前可能會有以下情形：

1. 資料格式不一致 2. 資料不完整 3. 資料無效 4. 資料重複
機器就像人一樣，吃下營養的食物就健康茁壯，我們越是給它精確有效的資料，機器就能做更有效的學習。因此，在蒐集資料之後，下一部便是整理資料，以避免所謂「dirty data」^註影響到資料系統的準確性與完整性。

舉例來說，在2018年時，台灣人工智慧實驗室與台北醫學大學附設醫院合作，使用病患的資料進行分析，可預測病患在四小時後併發敗血症的機率。這AI敗血症預測系統是擷取病患生理數據，再將數據使用演算法整合、分析研發而成。(資料來源：<https://www.tmuh.org.tw/media/2018>)。準確率的提升與否與數據是否有效，有很大的關係，如果沒有乾淨或有效的資料，系統無法發揮最大的效能。



圖2-4 丟掉dirty data才不會影響資料處理的結果



神救援

Dirty Data (髒數據)：Dirty Data 是指不完整、或包含錯誤訊息的數據資料。一般來說重複、不正確、不符合格式、或不完整的數據，都稱為髒數據；反之就是乾淨的數據。



阿圖出任務

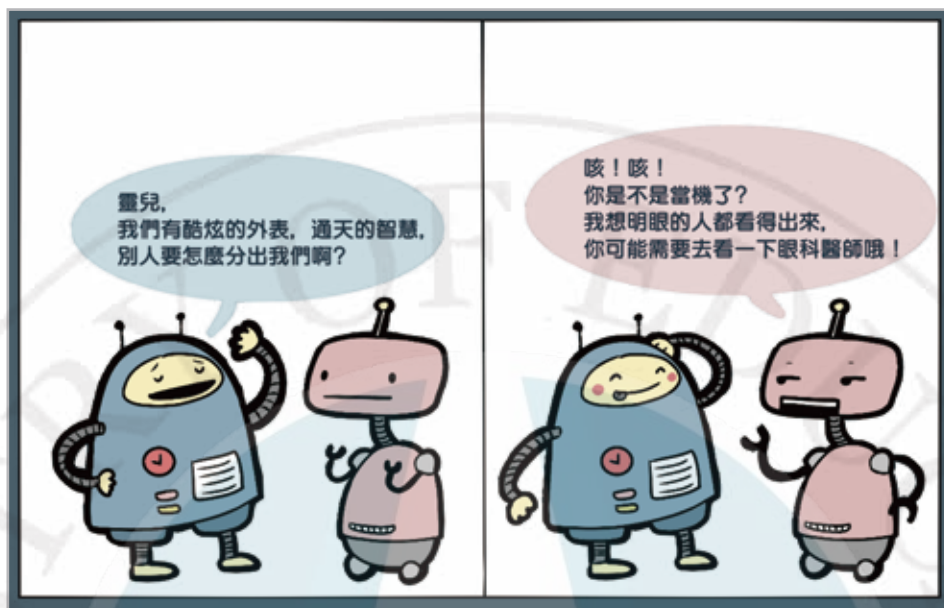


關卡三 • 去蕪存菁

以下資料是小叮從自己鉛筆盒中的文具量測的數據，為了之後的資料分析，我們必須先對資料做處理，請你針對文具可能的屬性，將下表中錯誤的、無用的資料直接用紅筆劃掉，並說明它可能錯誤的原因。

文具名稱	長度（單位：公分）	重量（單位：公克）	可能錯誤之原因
原子筆	14.5	3	
原子筆	14.2	10	
原子筆	14.2	12	
原子筆	1.3.7	9.8	
原子筆	13.9	10	
原子筆	142.7	12	
原子筆	14.1	98	
原子筆	11	10	
原子筆	145	11	
原子筆	14	8	
原子筆	14.5	9	

2-2 揪出同夥人



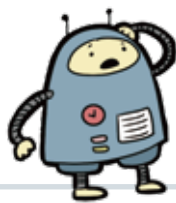
等級 練

你看過恐龍圖鑑嗎？恐龍種類這麼多，是如何分辨出暴龍、雷龍、還是迅猛龍？認識越多動物之後，你可以分辨出貓還是狗；聽到音樂前奏，就能分辨出是哪首歌、或者使用哪種樂器；想想看，有哪些辨別的地方呢？

- 1 看圖鑑辨識恐龍的方法有 → _____。
- 2 看圖片辨識豬或狗的方法有 → _____。
- 3 聽前奏辨別歌曲的方法有 → _____。
- 4 聽音樂辨別演奏樂器的方法有 → _____。



人工智慧系統如果要能正確分類，必須先進行**挑選特徵**，挑選的特徵越多，可以提供的資訊就會越多，但進行分析時會更耗時，也可能因為過多不相干的資訊導致分類能力下降。因此不論是挑選的特徵，或者是所挑選特徵的數量，兩者都很重要。



阿圖出任務

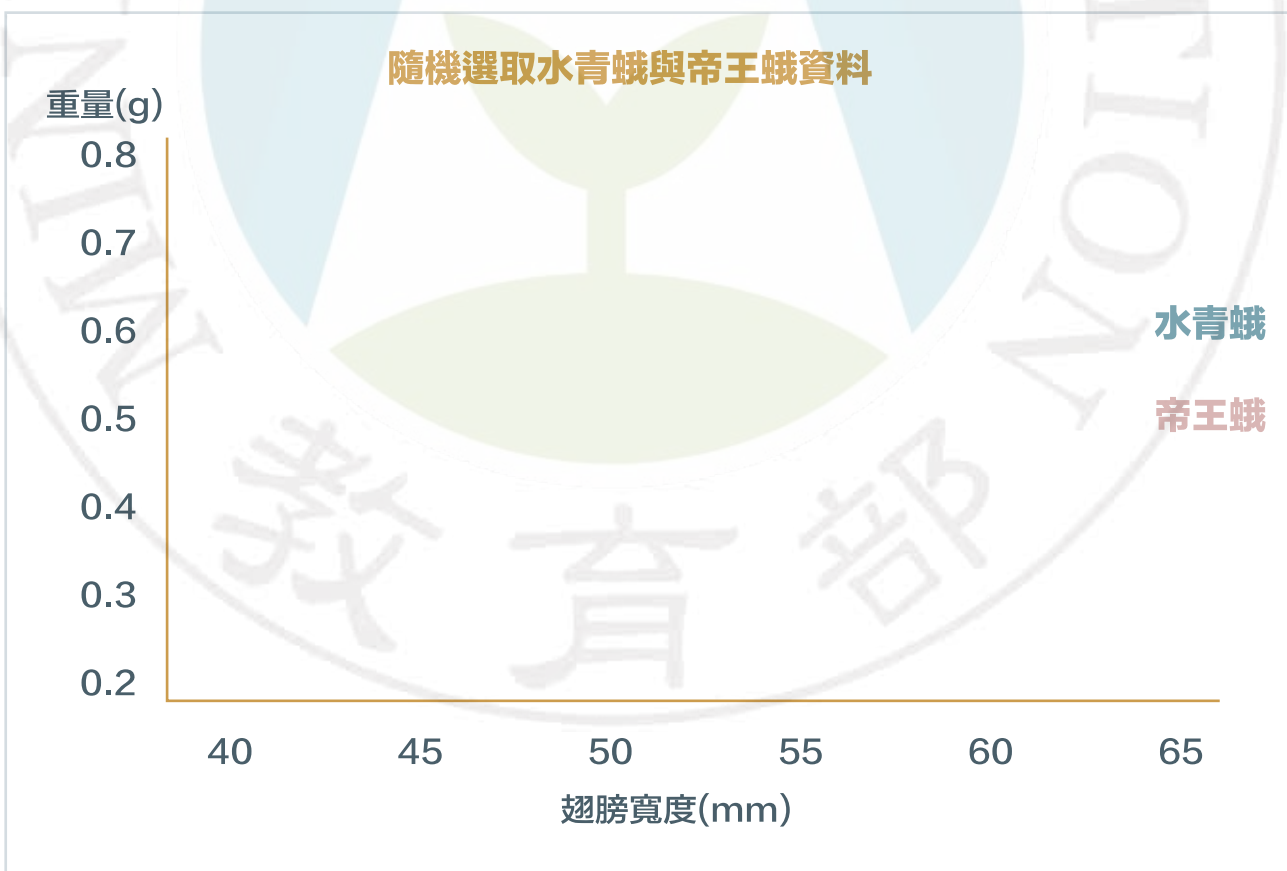


關卡四 • 似是「蛾」非

小叮的老師帶全班戶外教學，大家在野外看到成群飛舞的蛾，但是每一隻都長得好像喔~~要怎麼分辨牠們呢？於是老師跟大家介紹水青蛾與帝王蛾的外觀特性，並將16隻蛾（8隻水青蛾、8隻帝王蛾）分類好讓同學認識。

小叮心裡想：如果還有更多隻，我有沒有更好的辦法快速分辨呢？於是同學分別將每隻蛾的身長和重量標示出來，接下來找到的蛾就可以參考已知的16隻來做分辨了！

現在，你是不是可以幫忙把蛾的資料標示在坐標圖上，並且想一想，有甚麼好方法可以很快找到分類的規則呢？

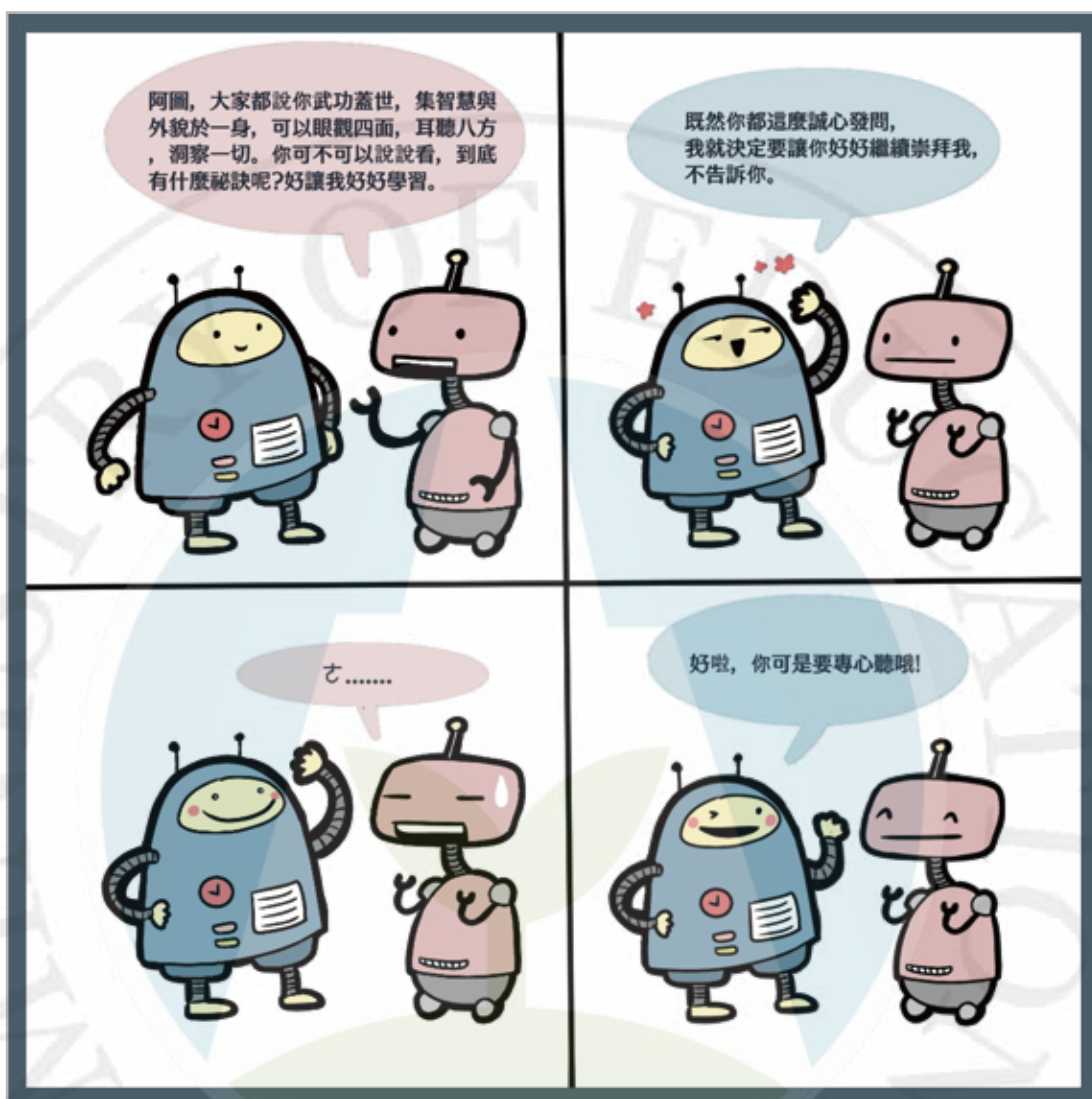




關卡五 ● 茫茫果海找同類

上一個任務你已經找到方法可以分類兩種相似的蛾，若現在有十種不同品種的芒果，我們可以用甚麼方法去分類它們呢？

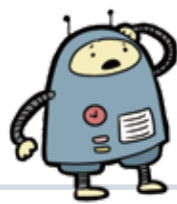
2-3 芝麻開門—快來探聽通關密碼



等級 練

小叮最近迷上籃球，但對於籃球的認識還不夠，想想看，有哪些通關密語可以讓小叮更認識籃球呢？

- 1 _____。
- 2 _____。
- 3 _____。
- 4 _____。
- 5 _____。



阿圖出任務



關卡六

AI密碼

要通往人工智慧大門，有十個通關密碼你一定要知道。當你收集好這十個通關密碼，就可以準備探索人工智慧的奧秘～

現在，請你就以下的說明，一一找出通關密碼吧！！

	通關密碼	密碼說明
1		只具有特定用途的人工智慧。
2		讓設備與設備之間可以互相溝通，或交換資料。 例如：智慧門鎖、用手機控制電扇開關等。
3		隨著時間不斷產生出來的大量、快速、且多樣性的資料。
4		為解決問題而設計出的一套流程與方法。
5		模擬人類大腦神經細胞網路的模型。
6		從大量資料庫中擷取出具有意義的資料之過程。
7		測試一台電腦是否可以與人對話，且不被分辨出它的機器身分。
8		機器為了達到最大或最佳效益，會依著環境而改變它的學習狀態。
9		人工智慧自我學習的技術，透過資料和經驗來學習，並找出規則，最後達到人工智慧的一種學習方式。
10		深度學習的分支，源自於類神經網路，它需要大量的樣本資料與強大的計算能力，例如AlphaGo。

密碼01.機器學習 (Machine Learning)

回顧我們在第一章所認識的人工智慧、機器學習與深度學習三者之間的關係。機器學習是人工智慧自我學習的技術，它是透過資料和經驗來學習，並找出規則，最後達到人工智慧的一種學習方式。機器學習的訓練方式必須要有大量的資料，從資料中訓練、學習，並且不斷修正。例如從消費者購買資料中得知其喜好、自動識別花朵種類等。

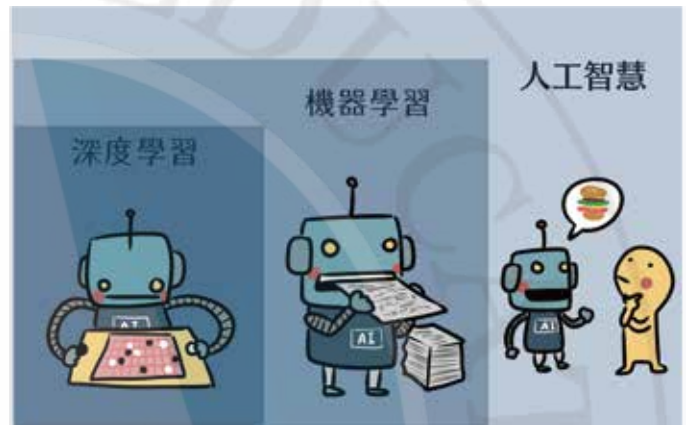


圖2-5 深度學習、機器學習與人工智慧三者之間的關係



圖2-6 類神經網路

密碼02.類神經網路 (Neural Network)

類神經網路是模擬大腦神經細胞網路的模型。這些網路以非常類似於人類神經系統和大腦的方法設計而成。透過不同學習階段，神經網路讓人工智慧能夠將資料分解為不同層級的數據，來解決複雜問題。

密碼03.深度學習 (Deep Learning)

深度學習是機器學習的分支，源自於類神經網路。深度學習主要有三個步驟：建構神經網路、設定學習目標、開始學習。進行深度學習必須有大量的樣本資料與強大的計算能力。以 AlphaGo 為例，研發團隊首先先建立多層次神經網路，之後輸入大量的棋譜資料，讓機器開始學習下棋的方法及各種應對，經過反覆且大量的訓練過程，機器便能根據對手所下的棋，找出最佳的回應方式。

第二章：從不同中發現相同－人工智慧的背景知識

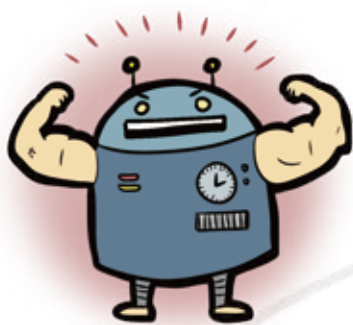


圖2-7 增強式學習會讓機器學習更加強大

密碼04.增強式學習 (Reinforcement Learning)

增強式學習是機器學習中的一個子領域，主要強調機器為了達到最大或者最佳效益，會依著環境而改變其學習狀態。

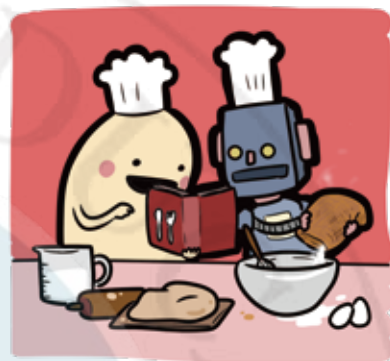


圖2-8 弱人工智慧是指只有部分特定用途的人工智慧

密碼05. 弱人工智慧與強人工智慧

人工智慧的目的是為了使機器具有人一樣的智慧，而像人類一樣具有廣泛智慧的人工智慧稱之為「強人工智慧」；反之，只具有特定用途的人工智慧稱之為「弱人工智慧」，例如：影像辨識、車輛偵測等。

密碼06.演算法 (Algorithm)

演算法是為了解決問題，所設計出一套解決問題的程序與方法。依照問題及需求，規劃出一套解決問題的流程與方法，再依照流程開始撰寫或設計程式。一樣的問題可以有多套演算法，但演算法設計的好壞會直接影響解決問題的效能。



圖2-9 演算法是解決問題之前所設計出來的一套程序與方法



圖2-10 圖靈測試示意圖

密碼07.圖靈測試 (Turing Test)

圖靈測試是由英國科學家艾倫圖靈於1950年提出的理論，他提出如果機器能夠透過模仿進而產生思考且與人展開對話，而不被分辨出他的機器身份，就表示這台機器具有人工智慧，我們稱這種測試為圖靈測試。



圖2-11 多樣、大量且快速的大數據

密碼08.大數據 (Big Data)

大數據又稱為「巨量資料」，大數據是指大量 (Volume)、快速 (Velocity) 和多樣性 (Variety) 的資料，隨著時間不斷增加的資料量總稱。大數據的特性是來源多元、種類多，更新速度快。以Facebook來說，每分每秒的貼文、回覆、照片、按「讚」數；Google搜尋引擎關鍵字搜尋次數、網站登錄紀錄，這些都屬於大數據的範疇。越來越多人利用大數據進行各種分析研究，並將其擴展運用。例如：分析消費者喜好，設計行銷策略等。

密碼09.資料探勘 (Data Mining)

資料探勘是從大量資料中擷取出具有意義之資料的過程，又稱資料採礦。例如某購物網站利用顧客資訊 (年齡、性別、學歷等) 及顧客消費紀錄 (購買種類、款式、價格、頻率等)，分析客戶消費模式，並依此結果將客戶分類，針對不同客戶群做出不同的行銷策略，這就是一種資料探勘。



圖2-12 資料探勘 (資料採礦)

密碼10.IoT 物聯網

IoT 是Internet of Things 的縮寫，中文名稱是物聯網。「物聯網」顧名思義是將東西串成一個網路，實質意義是使設備與設備之間可以互相連接，並做資料交換或溝通。人們的生活方式，將因為物聯網而有了改變。物聯網的應用非常廣泛，例如：智慧門鎖、空氣品質監測、土壤監測、智慧家電等，物聯網已經在不知不覺中慢慢改變我們的生活，也更方便我們的生活。

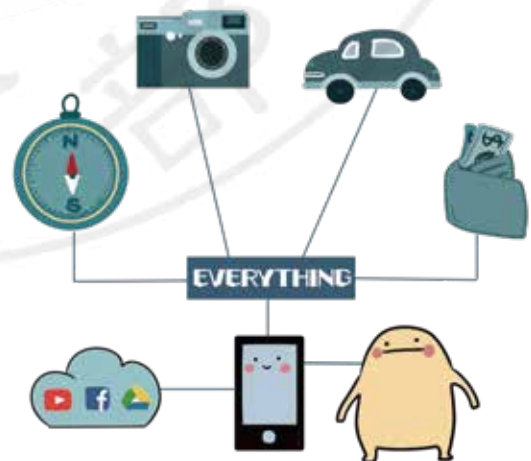
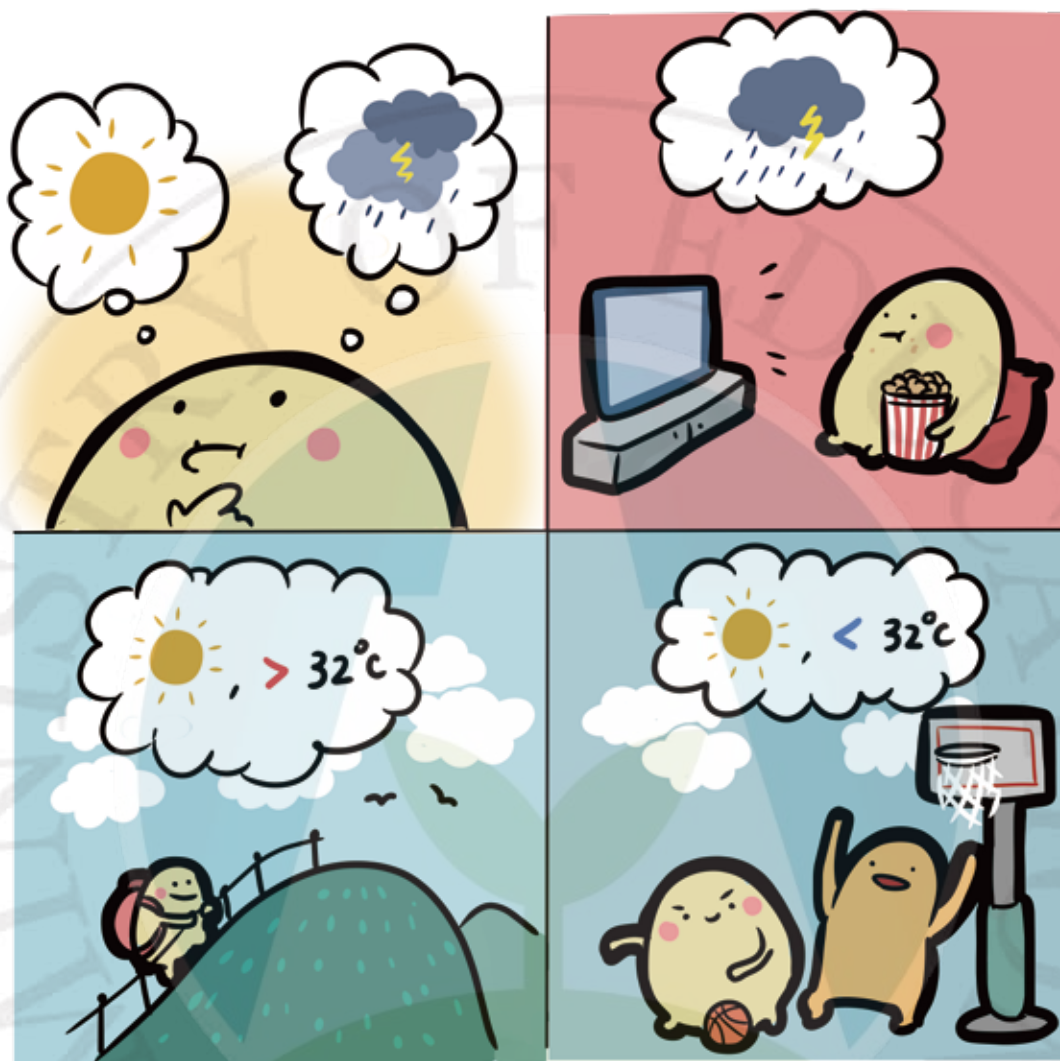


圖2-13 物聯網示意圖

3-1 「向左走向右走」－決策樹(Decision tree)



練等級

想想看，剛考完試的周末，你會想要做甚麼事情放鬆一下呢？
除了天氣會影響你的選擇之外，還會受到甚麼因素影響呢？

想做的事情是 ➡ _____ 、 _____ 、 _____ 、 _____ 。

影響的因素是 ➡ _____ 、 _____ 、 _____ 、 _____ 。



阿圖出任務



關卡一 ● 超強讀心術

以下的任務必須用到「決策樹」，請同學先去「檢裝備」（10分鐘），再試著幫小叮把週六的計畫填入下面決策樹中。

星期六下午，小叮正在思考要不要出門，小叮想看天氣狀況來決定：

- 如果外面下雨，選擇在家追劇。
- 如果沒有下雨，且外面天氣很熱（氣溫大於 32°C ），去爬山好了～
- 如果沒有下雨，外面是個舒服的天氣（氣溫小於 32°C ），找麻吉去打球。

小叮的計畫

Q1：提示-考慮天氣

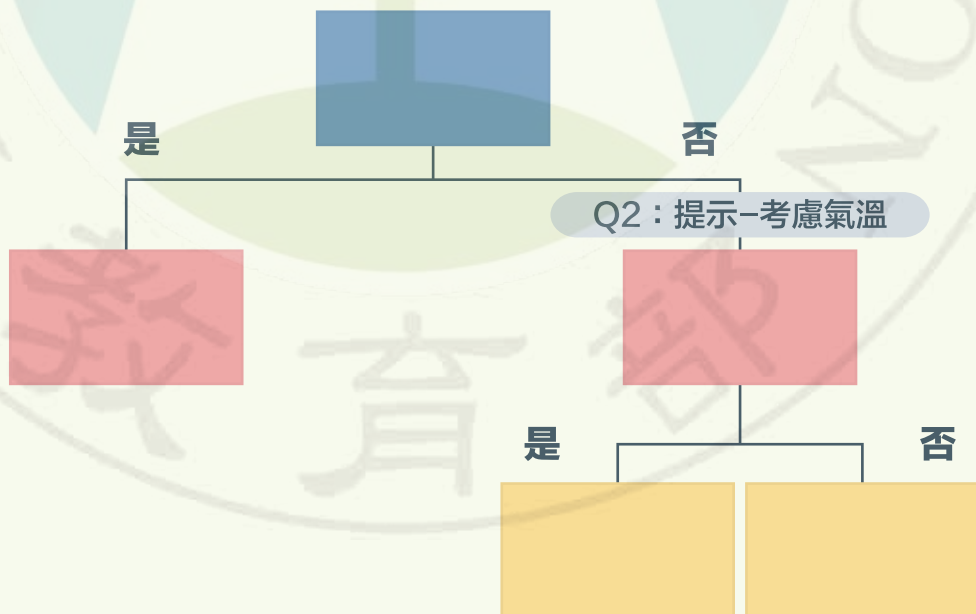


圖3-1 小叮週六計畫決策樹



監督式學習 (Supervised Learning)

為了方便理解，我們可以把監督式學習視為「有老師指導」的學習方法，因為老師擁有正確答案(在監督式學習裡這個答案稱之為標記)，在訓練的過程中，我們會告訴機器答案，也就是提供每個訓練樣本對應的標記資訊。例如，我們給機器看已經事先標記的1000張貓或狗之照片，機器經過訓練後，我們拿一張新的照片詢問它照片中的動物是貓還是狗，這過程就是監督式學習。在這一章中，同學們會學到監督式學習中的決策樹、感知器以及線性迴歸等方法的運作原理。



圖3-2 監督式學習

檢裝備

「決策樹」在人工智慧領域中，是一種很常見的資料分類方法，也是一種效率很高的監督式機器學習模型。它是類似一棵樹的架構，我們稱它為「樹狀結構」。如果我們要將一個未知類別的資料用決策樹來做分類，則高度越低的樹，因為判斷的次數較少，效率越高。一個決策樹包含了根部節點、決策節點及葉節點，呈現方式如圖3-3所示。

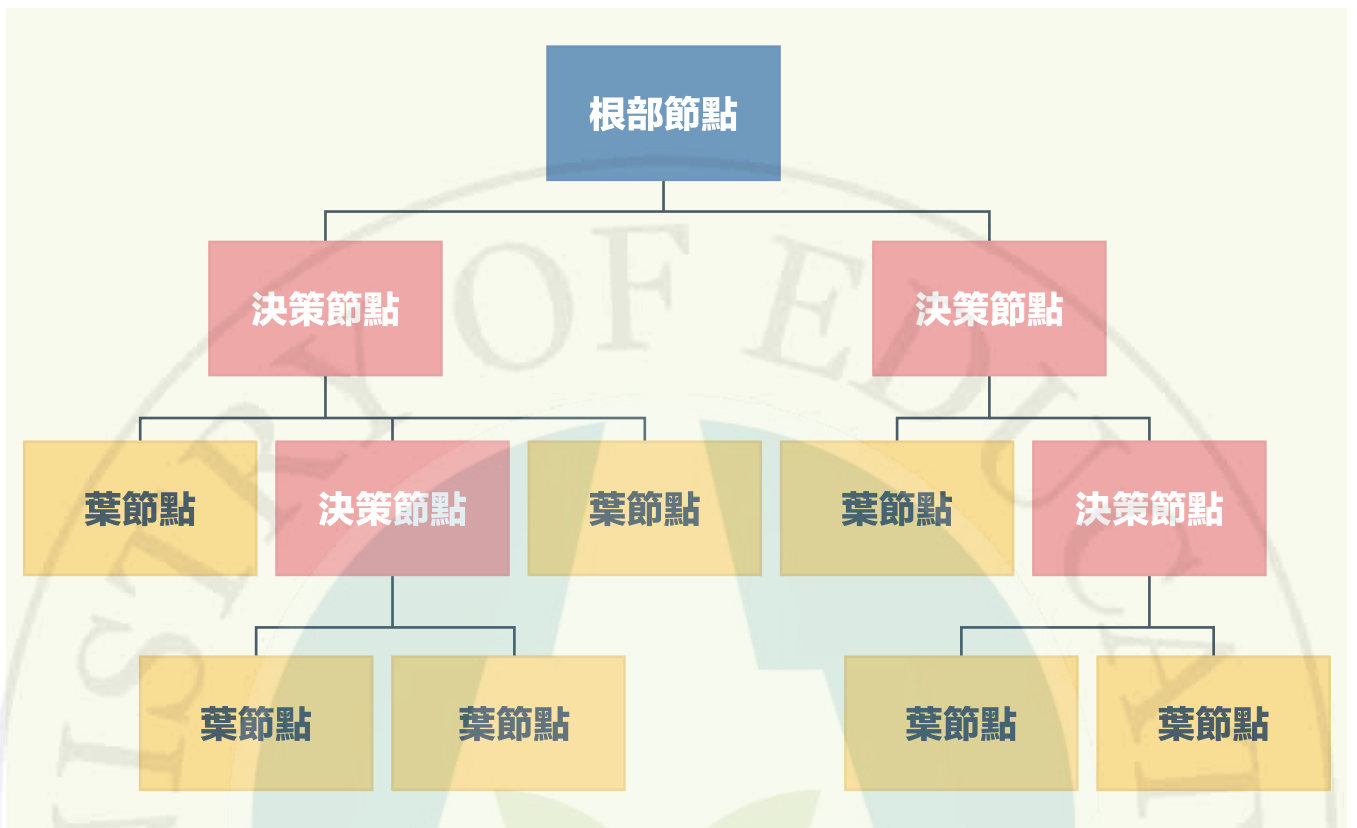
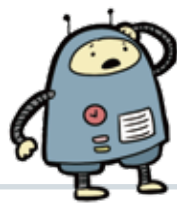


圖3-3 決策樹架構圖（※各節點依其特性以顏色區分）

每一棵決策樹只能有一個根部節點，但可以有多個決策節點及多個葉節點。

各節點說明如下：

- **根部節點**：樹的第一層，也是最開始的問題點。
- **決策節點**：決策過程的分歧點，需選擇一個問題來決定進入下一層的節點。
- **葉節點**：或稱終端節點，是決策或分類的結果。



阿圖出任務



關卡二 • 內心小劇場

當你在做抉擇的時候，心裡一定有更多影響你做決策的因素，請任意選擇一個問題，試著在下圖方框內，把問題與決策繪製成一棵決策樹，並與同學和老師分享。

例如：

- 1 今天早餐該吃甚麼？
(中式或西式?甜的或鹹的?冰的或熱的?)
- 2 好朋友生日快到了，該送他什麼？
(朋友的性別?喜好?顏色?價錢?類型?)
- 3 有人約我看電影，我要不要去呢？
(朋友請客嗎?朋友個性如何?兩人交情如何?)



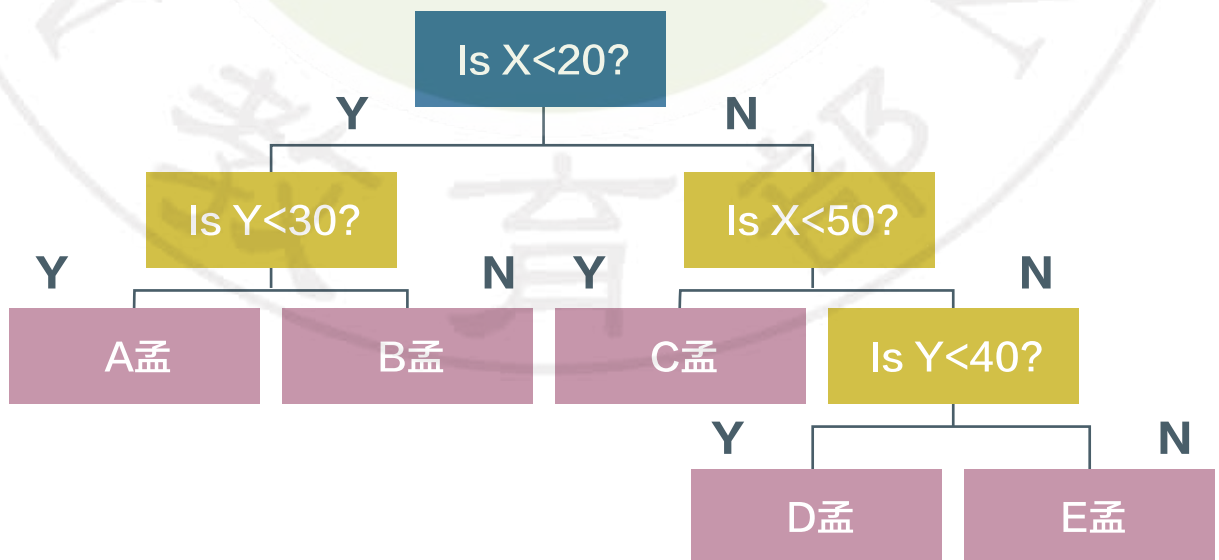


關卡三 • 決策樹達人

1 要完成此關卡必須善用提示，請同學先到「神救援」充電一下。（10分鐘）想想看，現有A、B、C三件文具，請你參考文具分類決策樹，依照下表A、B、C的特徵，預測它們分別屬於哪一種文具。請將答案填入空格中。

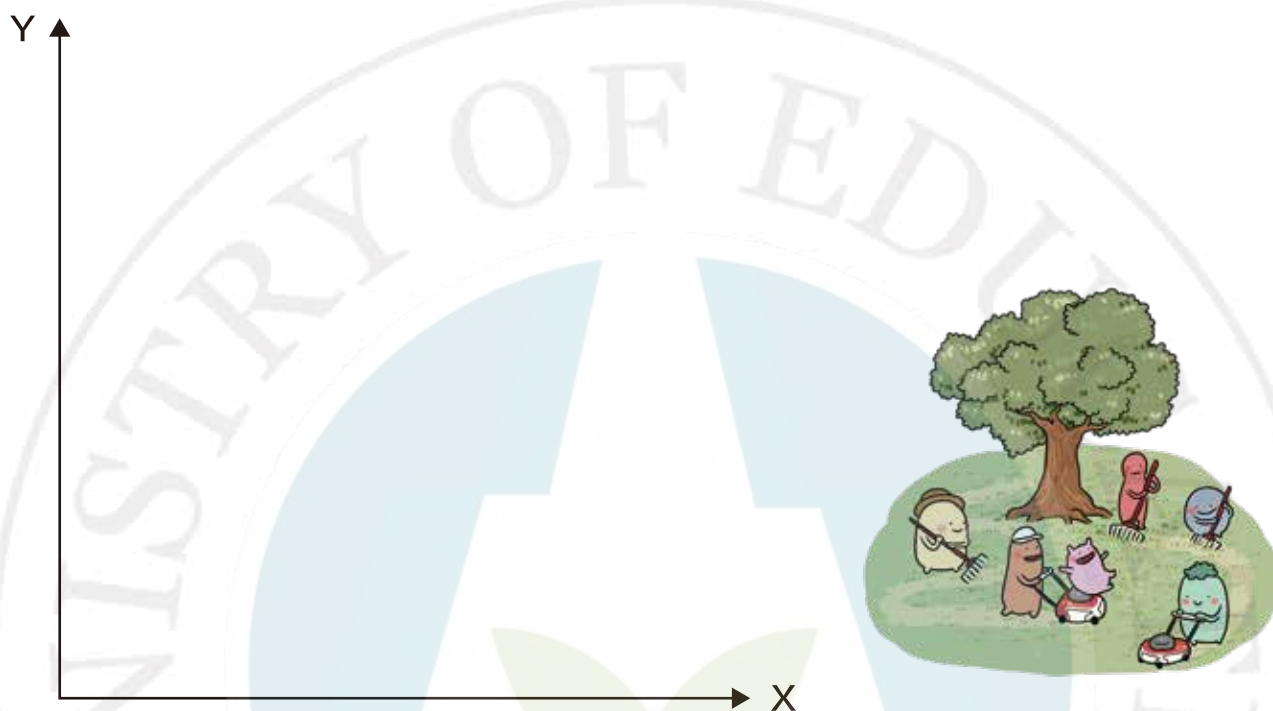
代號	長度（公分）	重量	文具類別
A	12	10	
B	16.5	15	
C	13	20	

2 A孟、B孟、C孟、D孟與E孟五兄弟是小叮的鄰居，他們的除草區域分配如下圖的決策樹，請試著在XY座標上標記出五兄弟各自負責的區域。



第三章：從樣本中大膽預言－監督式學習

請利用X Y座標圖標示出五兄弟各自的區域。



3

(承上題) 院子裡有個郵筒與聖誕樹，聖誕節快到了，媽媽說誰的區域裏面有郵筒或聖誕樹就要分別負責寄信或裝飾聖誕樹。請利用除草區域決策樹或XY座標圖，判斷誰該負責寄信、收信？誰該負責佈置聖誕樹？

	位置	負責人
郵筒	$(X,Y) = (15,50)$	
聖誕樹	$(X,Y) = (60,35)$	



神救援

決策樹常用在分類資料上，連續型的資料。下表是小叮班上同學將自己常用的文具（包含直尺、橡皮擦、原子筆和修正帶等四項）進行長度和重量的測量後，所得之數據資料，如下表3-1~表3-4所示。

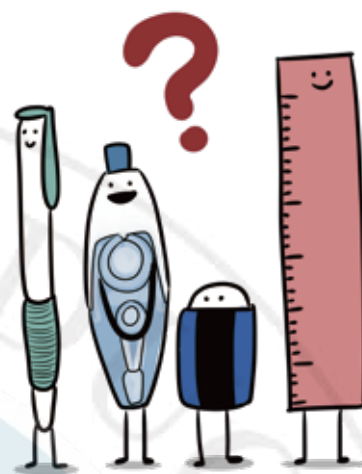


圖3-4 鉛筆盒中的文具四寶

表3-1

文具名稱	長度 (單位：公分)	重量 (單位：公克)
直尺	17.0	1.0
直尺	16.0	19.0
直尺	15.8	9.0
直尺	15.0	9.3
直尺	15.0	5.0
直尺	15.0	9.0
直尺	17.3	9.3
直尺	15	8.7
直尺	17	8.0
直尺	15	10.0
直尺	16.0	9.0

表3-2

文具名稱	長度 (單位：公分)	重量 (單位：公克)
修正帶	11.2	14.0
修正帶	11.0	20.0
修正帶	11.1	19.0
修正帶	11.0	17.6
修正帶	11.5	20.0
修正帶	11.0	20.0
修正帶	11.1	17.6
修正帶	11.0	20.3
修正帶	10.6	19.0
修正帶	11.0	19.0
修正帶	11.0	20.0

第三章：從樣本中大膽預言－監督式學習

表3-3

文具名稱	長度 (單位：公分)	重量 (單位：公克)
原子筆	14.5	4.0
原子筆	14.2	10.0
原子筆	14.2	12.0
原子筆	13.7	9.8
原子筆	13.9	10.0
原子筆	14.3	12.0
原子筆	14.1	9.8
原子筆	11.0	10.0
原子筆	14.5	11.0
原子筆	14.0	8.0
原子筆	14.5	9.0

表3-4

文具名稱	長度 (單位：公分)	重量 (單位：公克)
橡皮擦	6.4	16.0
橡皮擦	13.0	8.0
橡皮擦	2.8	10.0
橡皮擦	3.5	21.3
橡皮擦	9.5	4.0
橡皮擦	4.5	11.0
橡皮擦	6.6	21.3
橡皮擦	5.4	4.5
橡皮擦	4.2	10.0
橡皮擦	5.0	20.0
橡皮擦	1.0	12.0

接著將所有長度特徵設為X（橫）座標；重量特徵設為Y（縱）座標，在XY座標圖上繪製出來如下圖：

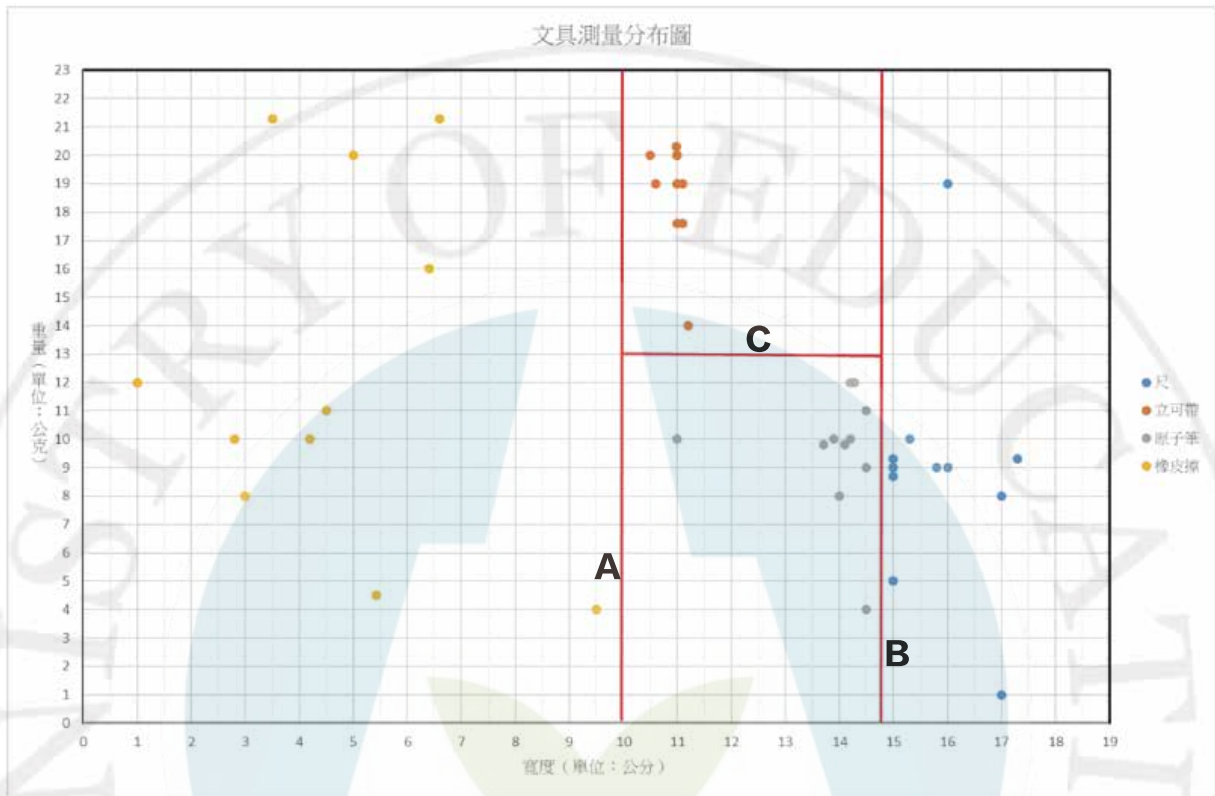


圖3-5 四種文具於二維座標上的分類結果

如果依據四種文具在坐標平面上的分佈，可以很容易找出彼此之間分類的界線，依序為A→B→C。先使用長度特徵找到A、B，接著使用重量特徵找到C，這時就可以畫成決策樹（如圖3-6）來表示，請你將文具分類決策樹的空格填滿。

文具的分類

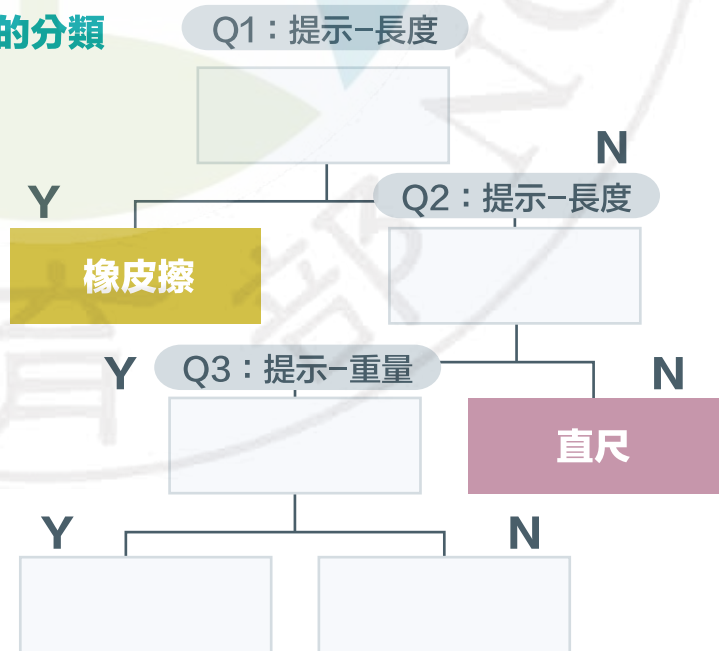


圖3-6文具分類決策樹

第三章：從樣本中大膽預言－監督式學習

繪製好的決策樹可以用來判斷未知的文具類別。例如：靈兒有一個文具，長度大約是18公分，重量是20公克，利用決策樹，我們可以大膽預言這個文具是「直尺」。

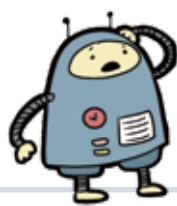
特徵的選擇及順序，在決策樹分類上扮演重要的角色，如果沒有慎選分類特徵的順序，可能會使建立的決策樹更加複雜，或者無法畫出一棵好的決策樹。



想一想

如果文具分類特徵選擇時，先選擇重量，再選擇長度，結果會是如何？請分享你的想法，並說明為何？





阿圖出任務



關卡四 • 撲克牌分類

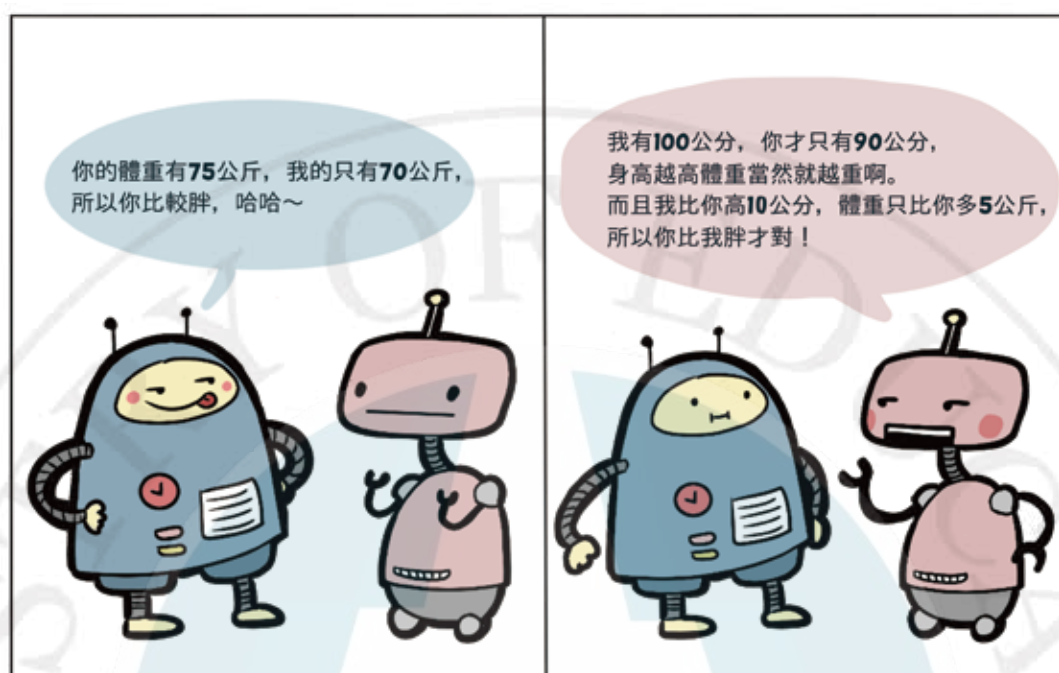
一副撲克牌有52張，如果我們想將撲克牌依照不同特性分類的話，有哪些分類方法呢？請同學四人一組，集思廣益，至少將撲克牌分成8類，有沒有甚麼方法呢？

提示：可以先分成兩大類，再將每一類細分，依此類推。

同學也可以去撿裝備找靈感喔！

組員	
組別：第	第一次分類：
組	第二次分類：
	第三次分類：
	Q：如果是黑色梅花5，會被分類到哪一類呢？

3-2 「誰上線？誰離線？」－線性回歸



各位同學，想想看，你們覺得是阿圖的話有道理，還是靈兒的話有道理？接下來我們將介紹，一個可以用來分析身高和體重兩個變數的規律並推測給定身高是否能推測出體重的方法。



想想看，請輸入詞語間的關係，並和同學比較是否有看法不同的地方？

比較項目	食量 (範例)	人緣	體重	歌喉	廚藝	外語能力
職業						
運動員	大					
模特兒	小					
電影明星	小					
廚師	大					



關卡五 • 體適能可以預測嗎？

小叮上體育課時，老師說要測量同學的體適能，小叮有個疑惑，是不是身高高的人天生跑得快？是不是比較瘦的人在體適能的表現會比較好？於是做完測驗之後，小叮想要試著從中找到答案……我們一起來幫小叮吧！

要完成此關卡必須善用提示，請同學先到「神救援」充電一下。（10分鐘）

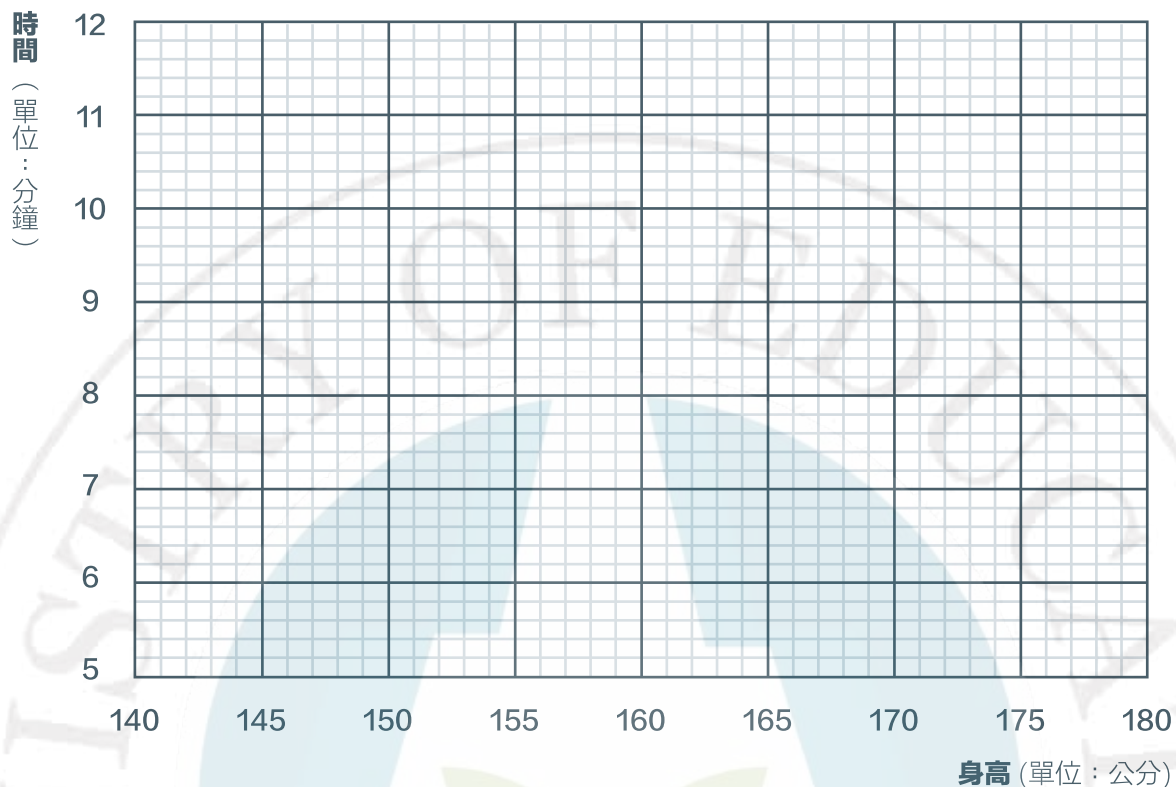
下表是10位同學跑1600公尺的跑步時間。如果想知道越高的同學是否可以跑得越快，請依序完成下列問題：



身高(公分)	1600公尺跑步時間(分)	身高(公分)	1600公尺跑步時間(分)
166.9	7.4	172.9	7.2
167.5	10.1	167.2	8.6
151.2	9.3	171.4	7.7
160.2	10.7	166	6.4
172.4	7.7	168	5.8

- 1 自變數是_____，依變數是_____。
- 2 請將上表資料之身高與1600公尺跑步時間依序描繪至圖上相對應的格點。
(注意各座標之相對位置，並正確描繪出)

身高與跑步（1600公尺）時間之回歸分析



- 3 依據各資料點，在圖上畫出一條最合適之趨勢線，也就是迴歸線（盡量使各點的誤差值達到最小）
- 4 請依據自己所繪製的迴歸線，使用自己的身高預估自己所需的時間。
 - 我的身高：_____公分，
 - 我預估自己的時間：_____分鐘

延伸思考：

如果有更多的資料，有沒有其他更快更準確的辦法可以找出最接近各點的迴歸線？



神救援

線性迴歸是假設自變數和依變數之間存在著線性關係。它是利用稱為線性迴歸方程式的最小平方法對自變數與依變數之間的關係進行建模分析。

自變數：這個變數不會被其他變數影響，它只會去影響別人，所以被認為是「原因」。



圖3-7 自變數



圖3-8 依變數

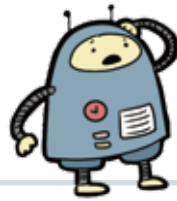
依變數：這個變數會被其他變數影響，因此被認為是「結果」。換句話說，我們可以將自變數視為「原因」，將依變數視為「結果」。

簡單來說，線性迴歸的意思就是「在眾多資料的分佈中，找出一條適切的直線，這條直線最能夠代表目前的資料走向」。

在座標中，給定兩個點，我們就可以帶入公式求得直線。線性迴歸是透過不斷調整直線方程式的各項係數，試圖找到一條最佳直線。這條直線可反映出自變數和依變數之間的關係，此時這條線稱之為「迴歸線」。



在迴歸線上，如果自變數越大，依變數也越大，此時我們會稱作兩者之間呈現「正相關」；反之，若自變數越大，依變數卻越小，此時我們會稱作兩者之間呈現「負相關」。如果兩個變數之間毫無關係，我們稱之為「零相關」。



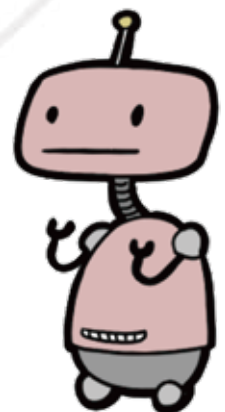
阿圖出任務



關卡六 • 靈兒幫幫忙

當我們需要處理的資料不再是5筆、10筆，這計算的工作勢必交由電腦來處理。在上個任務中我們用人腦去預測迴歸線，若資料量非常龐大時，計算的任務已不是人所能負荷。因此，我們需要電腦輔助。接下來，請跟著老師的指示，利用電腦幫我們找出迴歸線。

- 1 利用老師所提供之資料，使用Excel軟體之圖表繪製功能，繪製出XY散佈圖，顯示出所有資料點。
- 2 資料點皆繪製完成後，即可利用繪圖功能，繪出趨勢線。
- 3 試比較自己手繪估計的趨勢線，與電腦繪製出的趨勢線有甚麼差異？試著將電腦繪出之趨勢線，使用紅筆描繪在關卡六的練習中，並比較兩者的差異之處。





關卡七 • 隔空大比拚

- 請同學先做預測，再完成下表，並填入「+」（代表：正相關）、「-」（代表：負相關）與「0」（代表：零相關）。
- 使用Excel軟體，試分析以下任兩組資料之間，存在何種相關性？
（正相關、負相關、或零相關）



身高 (公分)	體重 (公斤)	BMI	坐姿體前彎 (公分)	立定跳遠 (公分)	仰臥起坐 (下)
166.9	64.5	23.2	25	230	39
155.3	54.0	22.4	25	150	25
167.5	57.5	20.5	15	140	30
164.9	38.9	14.3	17	150	30
151.2	53.3	23.3	25	150	37
159.8	54.4	21.3	23	180	40
160.2	50.0	19.5	25	170	40
165.1	50.4	18.5	13	180	37
172.4	58.7	19.7	35	230	48
161.5	55.5	21.3	18	140	21
159.2	66.5	26.2	14	160	32
172.9	59.3	19.8	24	180	42
139.2	40.3	20.8	25	100	33
167.2	43.2	15.5	13	160	31
171.4	63.0	21.4	12	260	42
166	54.9	19.9	19	190	35
168	51.4	18.2	29	200	52

第三章：從樣本中大膽預言－監督式學習

任務2：依據上個任務，將所得結果寫在下方，並繪圖（可螢幕截圖後印出貼上）。

1. 正相關：_____ 與 _____。

2. 負相關：_____ 與 _____。

3. 零相關：_____與_____。

延伸思考與分組討論：

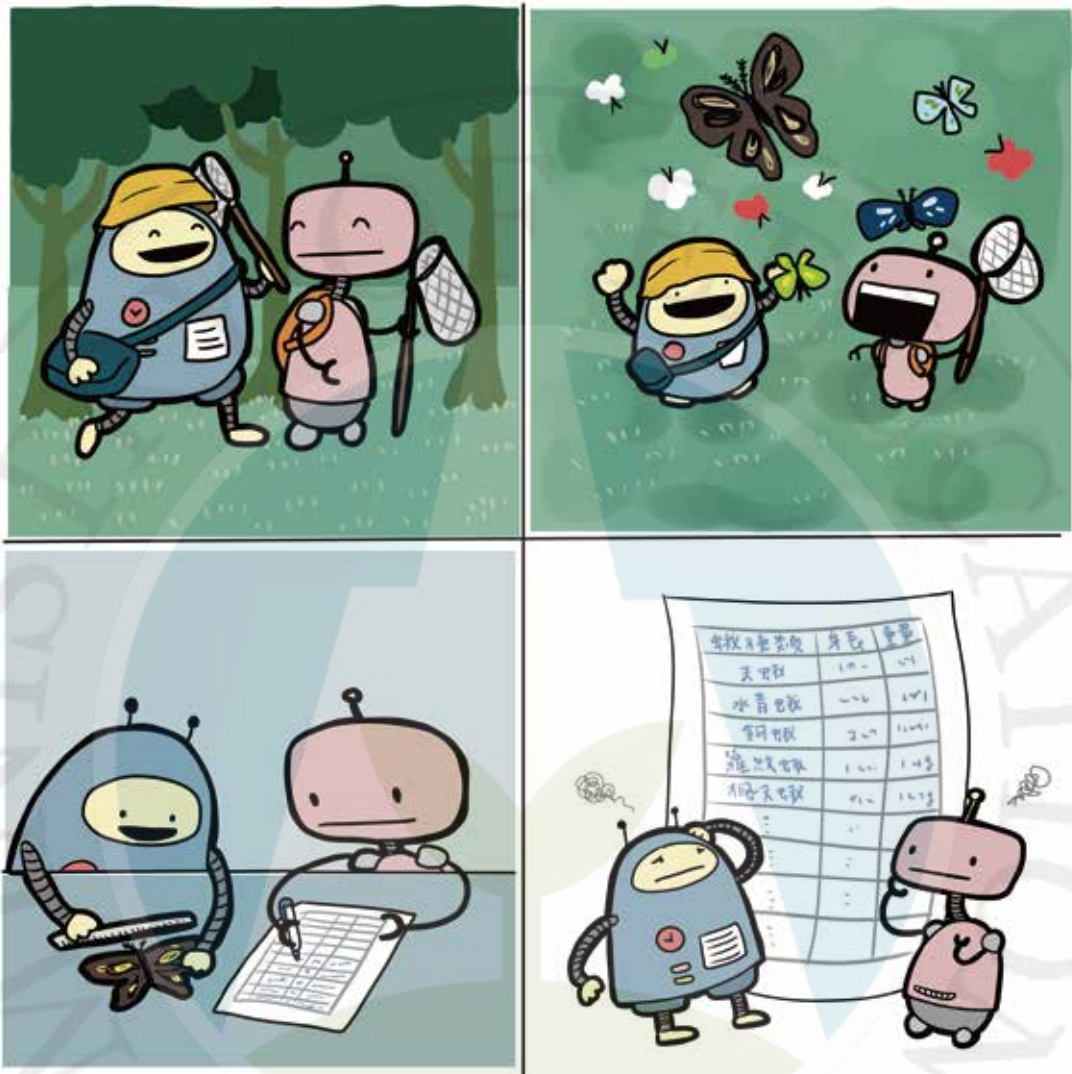
試想想，在我們的日常生活中、或校園生活中，還有沒有哪兩者資料之間關係是屬於正相關、負相關或零相關的呢？

○ **正相關：**

○ **負相關：**

○ **零相關：**

3-3 「線性?或非線性?」－感知器(Perceptron)



練等級

你知道如何辨別小小兵中的Bob（鮑伯）、Kevin（凱文）、Stuart（史都華）嗎？你知道如何辨別迪士尼系列中的白雪（白雪公主）、貝兒（美女與野獸）、仙杜瑞拉（仙履奇緣）嗎？想想看，有甚麼方法可以進行分辨呢？

我想要選擇 恐龍系列 公主系列 其他 _____

我的分辨方法是 _____



阿圖出任務



關卡八 • 此蛾非彼蛾

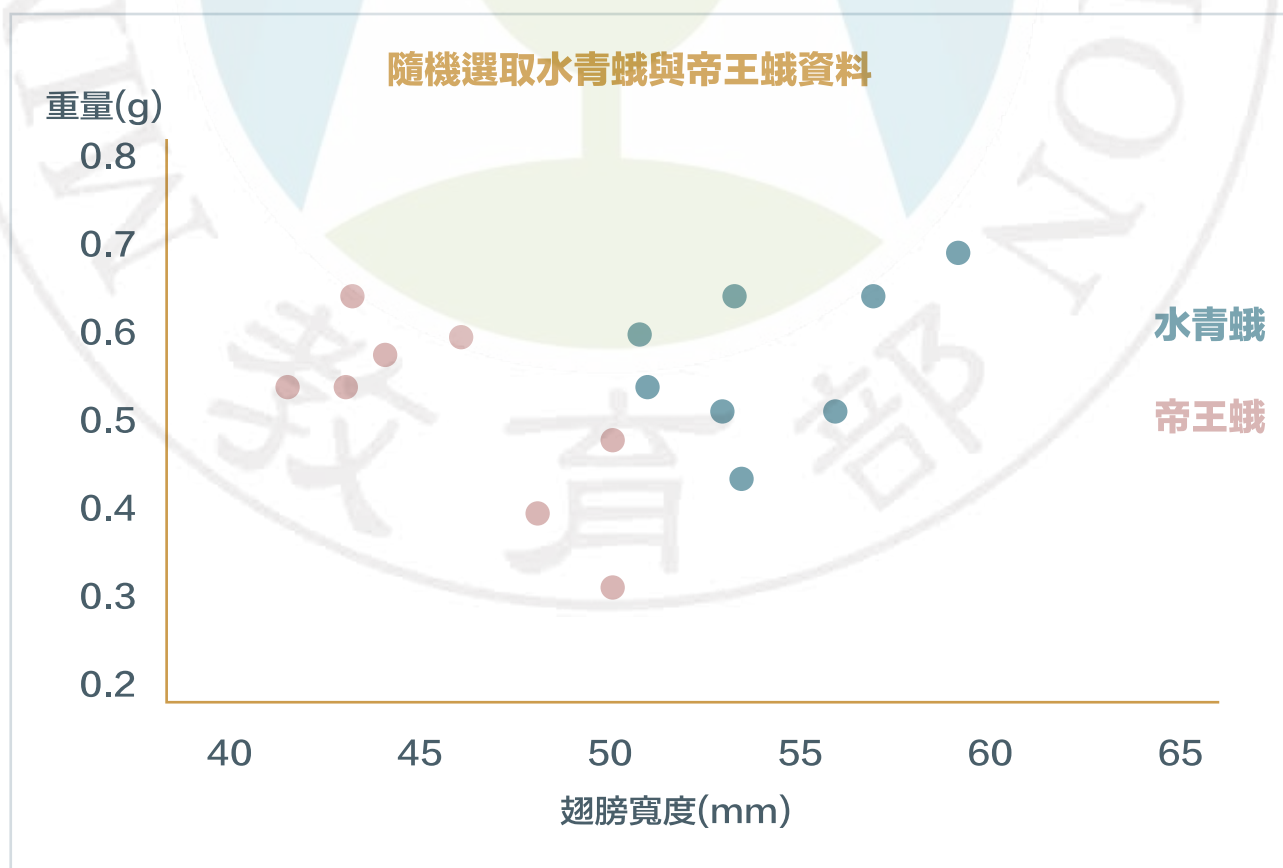
生物老師帶著小叮那班的同學到戶外進行考察，老師與同學收集了很多水青蛾與帝王蛾的身長與重量資料，你能不能依據大家收集到的資料，使用感知器的分類方法，將兩種蛾類分類呢？

說明：

感知器的演算法是一種訓練機器做線性分類的演算法，從被分類錯誤的資料中，慢慢調整感知器的參數，直到所有數據都被正確分類。

- 1 下圖中之資料點分別代表兩種蛾類(水青蛾、帝王蛾)之特徵分佈(翅膀寬度、重量)。請從下圖眾多資料中，使用感知器的分類方式，將所有資料正確分成兩類。

提示：將直線或繩子放在資料上方，漸漸移動其斜率及位置，直到找出能正確分類之直線後，使用膠帶將其固定。



第三章：從樣本中大膽預言－監督式學習

- 2 除了老師上課分享的範例之外，感知器是否還有其他的應用呢？或者你認為還有甚麼情況也適用感知器？

請與同學討論，並把你的想法或資料整理的結果寫下來。



感知器的理論是在1957年由美國神經學家 Frank Rosenblatt 所提出第一個神經網路模型，稱為感知器。感知器是用簡單的數學描述，模擬人類大腦中的神經訊息傳遞與觸發感知的行為。人類的神經元結構如圖3-9所示。

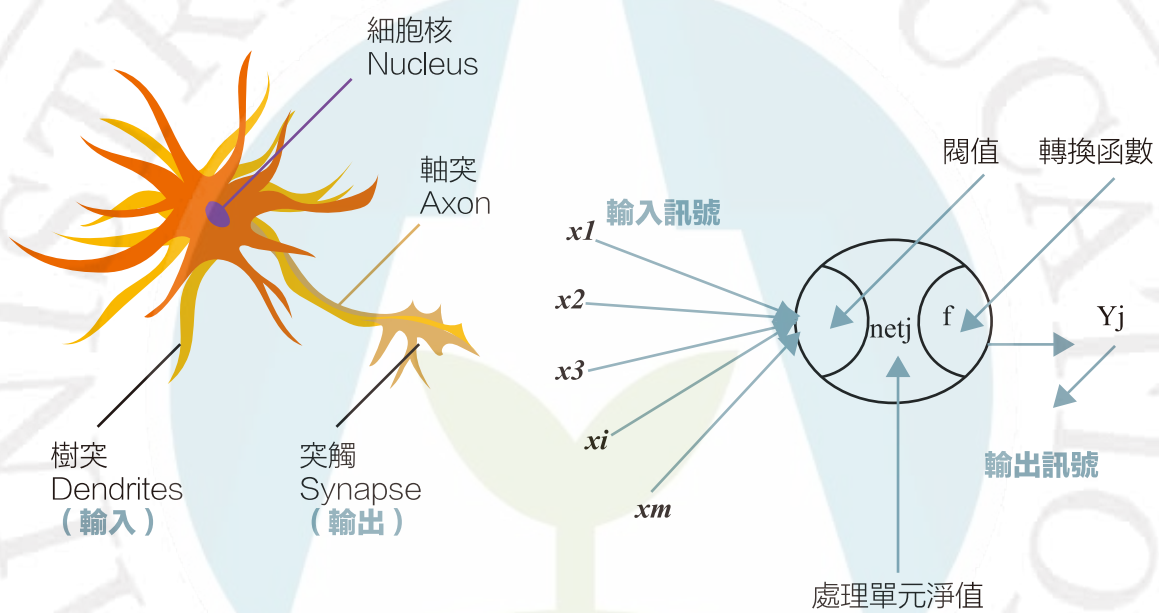


圖3-9 生物神經網路與類神經網路

在機器學習領域中，感知器是屬於監督式學習中的一種分類方法，主要用來做直線分類的二分類模型，也就是資料是在線性可分的狀態下才能正確被分類。所謂線性可分（如圖3-11），是指所有資料可以藉由一條適當的直線，將資料分成兩類，直線的兩側分屬不同類別。相反的，若資料無法線性分類（如圖3-12）則無法使用感知器來做為資料分類的工具。

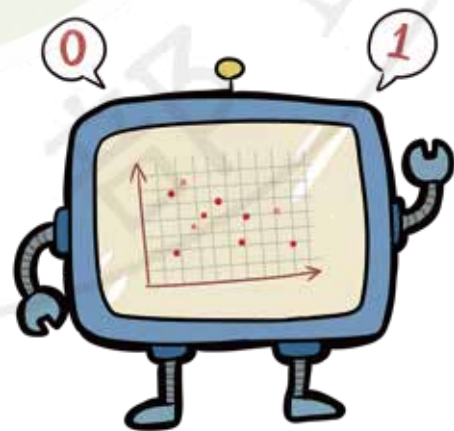


圖3-10 感知器

第三章：從樣本中大膽預言－監督式學習

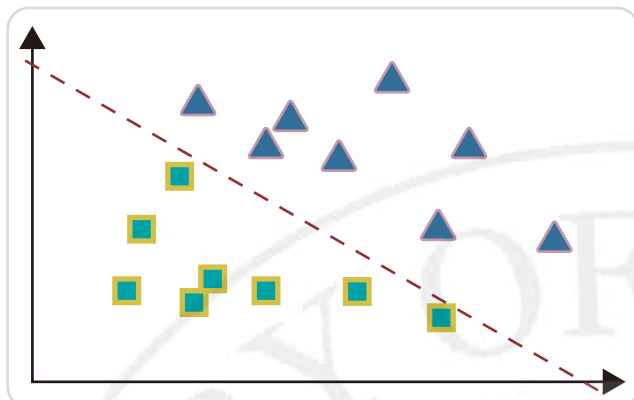


圖3-11 線性可分的資料

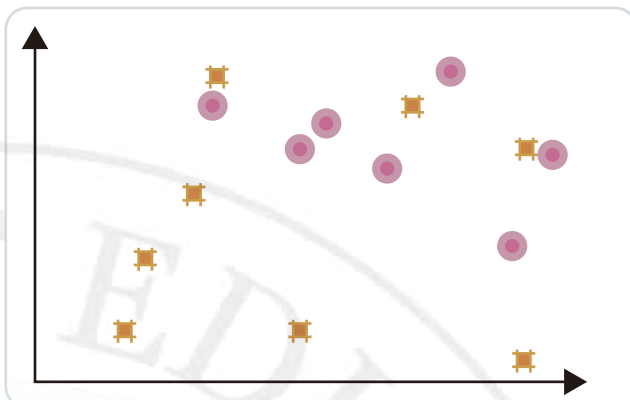


圖3-12 非線性可分的資料

感知器屬於單層神經網路，也是組成神經網路的最小架構。感知器的輸出，只有0或1兩種結果。舉例來說，感知器若使用在垃圾郵件分類上，常輸出為1代表該郵件被標記為垃圾郵件；若輸出為0，代表該郵件非垃圾郵件。

感知器的演算法是一種訓練機器做正確線性分類的演算法，從被分類錯誤的資料中，慢慢調整感知器的參數，直到所有數據都被正確分類。（分類原則如下所示）

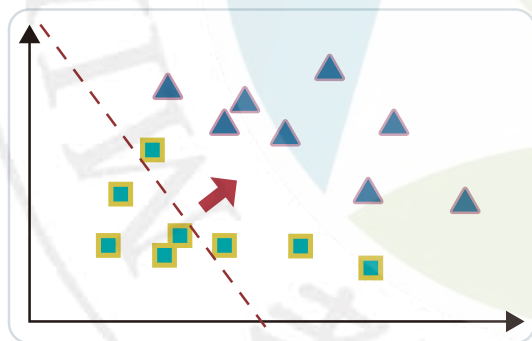


圖3-13-1 感知器分類過程

左圖資料中，有四個資料點被分類錯誤，修正時會將直線往減少分類錯誤的方向移動。（如紅色箭頭所示）

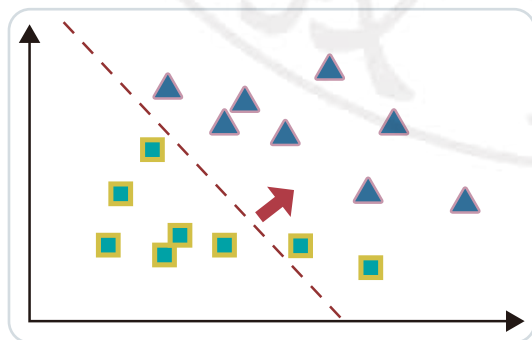
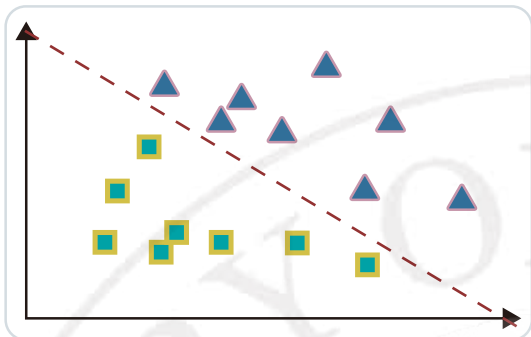


圖3-13-2 感知器分類過程

修正直線後，此步驟已減少至二個錯誤分類，此時直線將繼續往減少分類錯誤的方向移動。（如紅色箭頭所示）

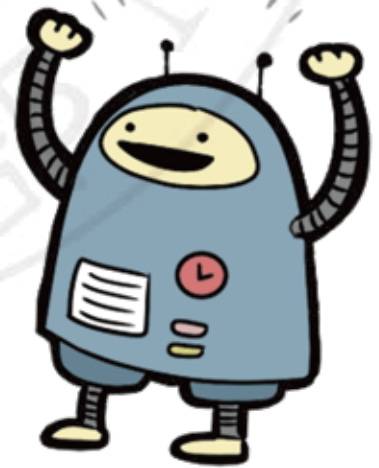


直線位置與斜率反覆修正，直到所有資料皆能被正確分類為止。

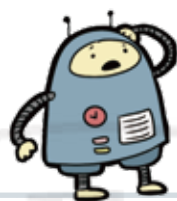
圖3-13-3 感知器分類過程

如上所示，感知器是一種監督式學習，在訓練過程中機器漸漸減少被錯誤分類的數量。

由於感知器只能用來做線性分類，但如果資料無法被線性分類，該怎麼辦呢？這時候就可以使用多層感知器，其具有非線性分類的的能力。



3-4 「看誰勢力大」-KNN分類器



阿圖出任務



關卡九 「投靠哪一隊？」

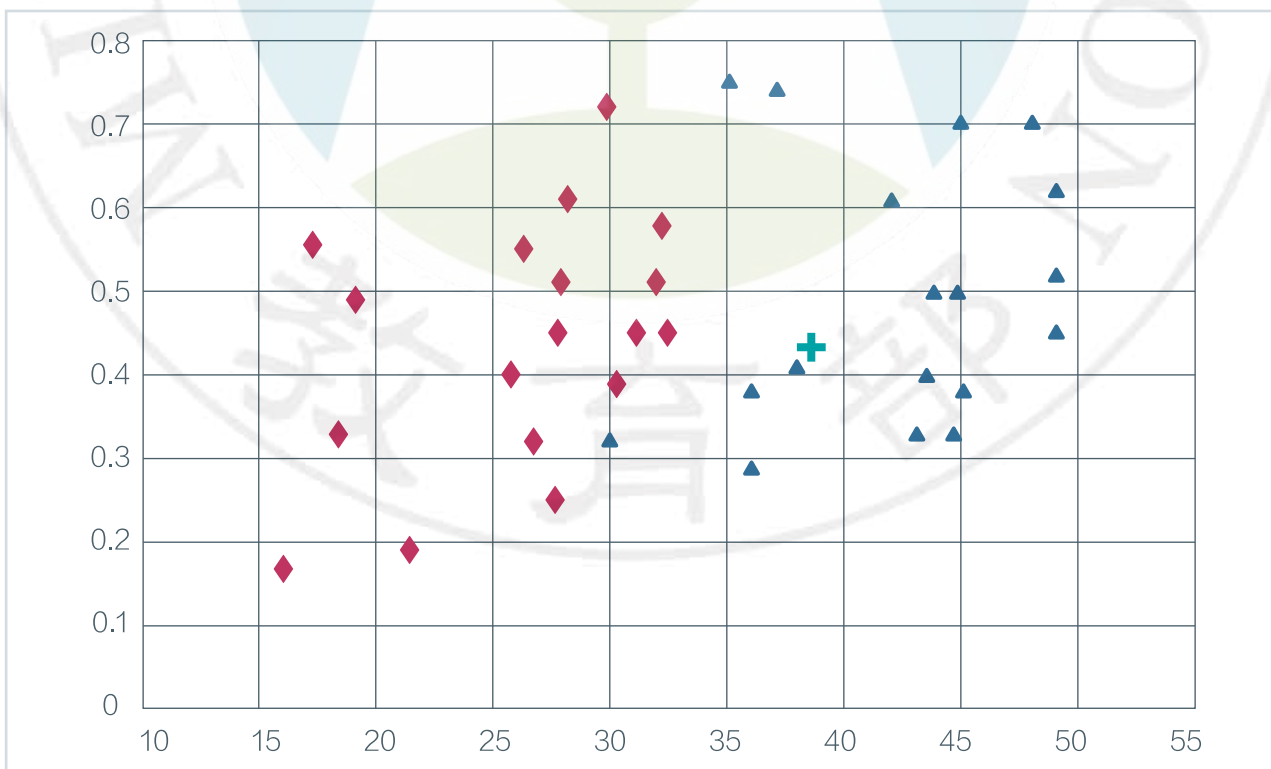
某天下午，小叮（代號+）到了漆彈場，發現場內已經有兩隊人馬，勢均力敵，小叮猶豫著應該加入哪一隊。請你幫小叮想想辦法，有甚麼方法可以決定小叮的所屬隊伍呢？

下列圖中之分類標示，以◆表示a隊，▲表示b隊。

1 使用直覺判斷，你認為小叮應該屬於那一隊？

Ans：_____隊（請填a或b）。

因為_____。



2 承上題，如果使用KNN分類器，根據不同的K值，小叮應該屬於哪一隊呢？
(記得去檢裝備喔！)

① K=3，小叮應屬於 _____ 隊。(請填 a 或 b)

② K=7，小叮應屬於 _____ 隊。(請填 a 或 b)

③ K=15，小叮應屬於 _____ 隊。(請填 a 或 b)

3 使用KNN分類器幫小叮歸隊之後，你是否有發現該演算法的優缺點呢？
請各寫出兩個於下表中。

優點	1.
	2.
缺點	1.
	2.

4 KNN鄰近演算法在人工智慧領域中，除了老師上課分享的範例之外，是否還有其他的應用。或者你認為還有甚麼情況也適用KNN鄰近演算法？請把你的想法或資料的結果寫下來，並與同學老師分享！



KNN分類器（以下簡稱KNN），是屬於監督式學習，也是機器學習最簡單的方法之一。在一群已標記好的資料中，若再加入新的資料，就可使用KNN來幫新的資料做分類。KNN的K是一個自訂參數，其原則找出與新資料點最近的K個資料，在這個K個資料中屬於哪個類別的資料最多，就把未知的資料歸為那個類別。

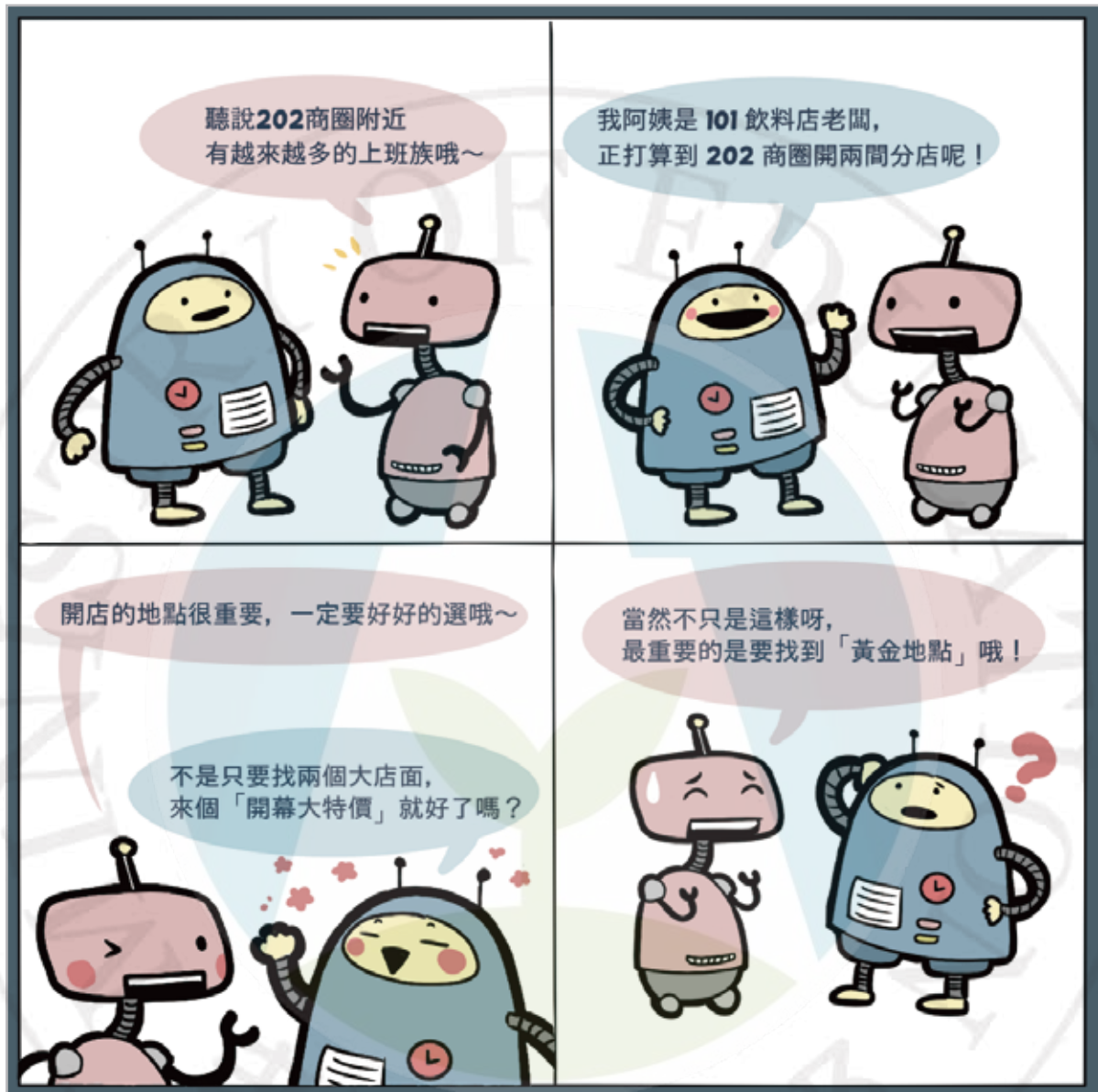
以下圖為例，已知之資料標記成兩類別，當加入新的資料(+)時，若 $K=5$ ，紅色虛線範圍內的藍色標記(▲)最多，此時新資料被歸類為藍色組；但是當 $K=15$ 時，綠色虛線範圍內的橘色標記(◆)最多，新資料則被歸類為橘色組。KNN就是這麼簡單的「看哪邊勢力大就靠那邊站」、「少數服從多數」的概念。



圖3-14 使用KNN分類示意圖

雖然KNN可以很簡單將未分類的資料作分類，但如何選擇適當的K值，顯得相對重要。較小的K值可以較快速得到結果，但通常比較容易出錯；較大的K值通常能夠得到比較佳的結果，但較耗時。一般來說，K值最好小於20且為奇數。雖然KNN可以很直接找到分類，但KNN有個缺點：使用KNN時，每次都必須計算欲分類的資料與全部資料的距離，才能得知新的資料是屬於哪一類，因此比較耗時。

4-1 「是我族或他類？」-K平均分群法

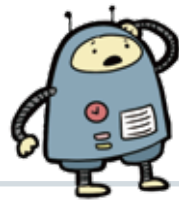


等級 練

想想看，除了賣各種好喝飲料的飲料店之外，還有甚麼店是你會經常光顧的，你覺得會賺錢的店？你覺得最適合開店的黃金地點會是在哪裡呢？

想要開的店是 ➡ 抓娃娃機店、_____、_____、_____。

黃金地點會是 ➡ 國中小附近、_____、_____、_____。



阿圖出任務



關卡一 ● 尋找中心

任務說明：（進行這個任務時，可以去「撿裝備」喔！）

請任選10張數字卡，依說明內容進行第一回合到第三~六回合的找中心活動，並進行分類任務。

第一回合

- 1 從數字卡中任取三個數字，將其數字分別填入 \triangle 內。
- 2 將其他牌卡依序分類至與自己最接近的數字組別。
- 3 分類好之後將所有牌卡的數字填入矩形方框內。
- 4 計算矩形方框內所有數字之平均值（四捨五入）並寫入下方 \circ 內。
- 5 再將 \circ 內之數字寫入第二回合之 \triangle 內。

第二回合

- 1 重新將所有牌卡分類至與自己最接近的數字組別，並填入矩形方框內。
- 2 計算矩形框內之所有數字之平均（四捨五入）並寫入下方 \circ 內。
- 3 將 \circ 內之數字填入第三回合之 \triangle 內。

第三回合（含）以後

- 1 重複以上操作步驟，直到同一輪的 \triangle 內數字與下方 \circ 內數字相同為止。

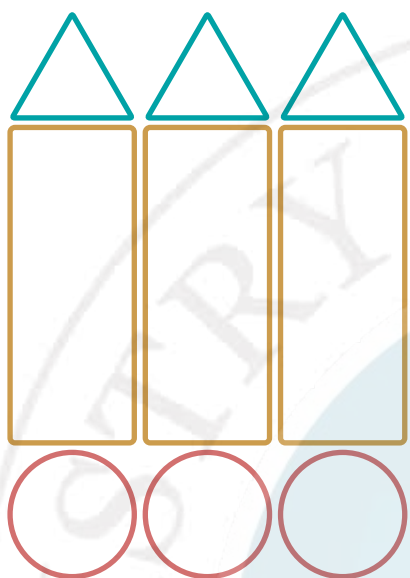
任務結果：

- 1 我的10個數字由小到大分別是_____。
（例如：1, 1, 2, 2, 2, 3, 5, 6……）
- 2 我的第一回合上方 \triangle 內之數字分別為：_____、_____、_____。
- 3 我的最後一回合上方 \triangle 內之數字分別為：_____、_____、_____。
- 4 我總共做了_____回合的分類，才讓 \triangle 內之數字才不再變化（意即收斂）。

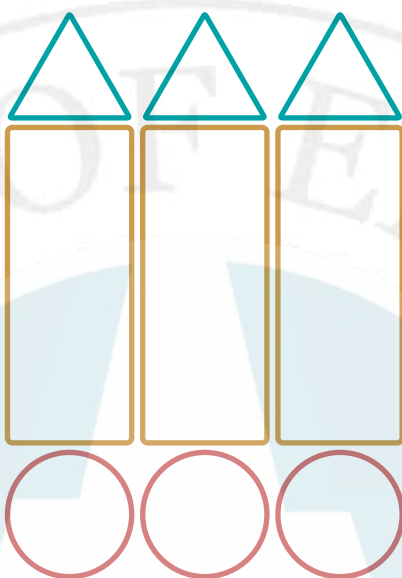


因數字卡皆為整數，這個任務我們四捨五入到整數，但實際上電腦運作是不需要四捨五入的哦！

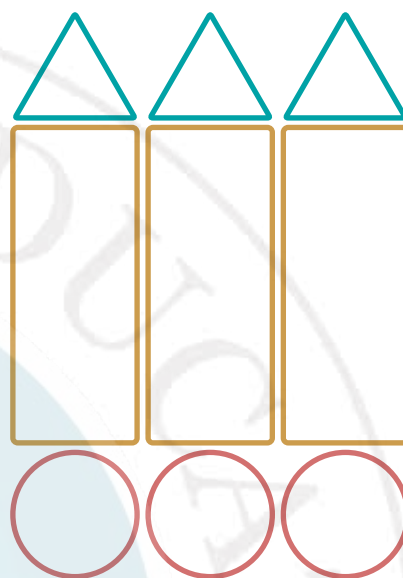
第一回合



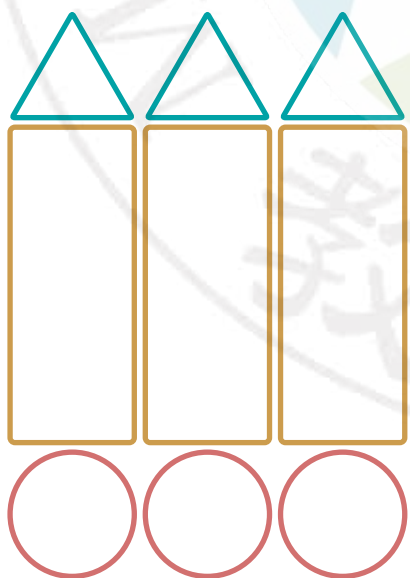
第二回合



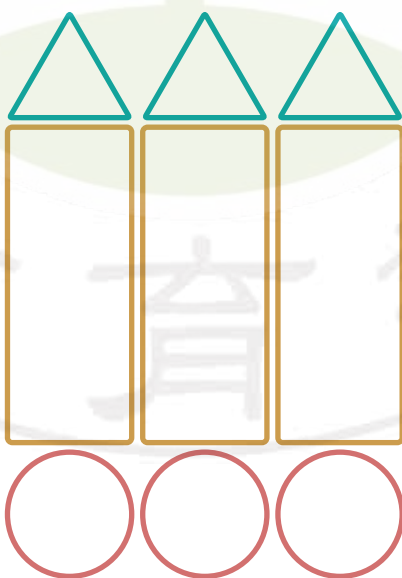
第三回合



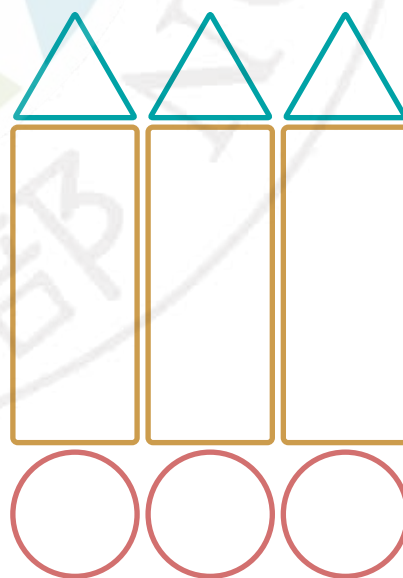
第四回合

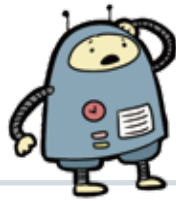


第五回合



第六回合





阿圖出任務



關卡二 • 披薩熱到家

小叮爸爸看準外送商機，打算在101大社區開設兩間『熱到家』Pizza店，目前小叮爸爸正在苦思分店地點。

請依照下面任務說明，幫小叮爸爸找出兩個最佳地點，並標示各房子應歸屬哪間分店的外送區域。

任務說明：

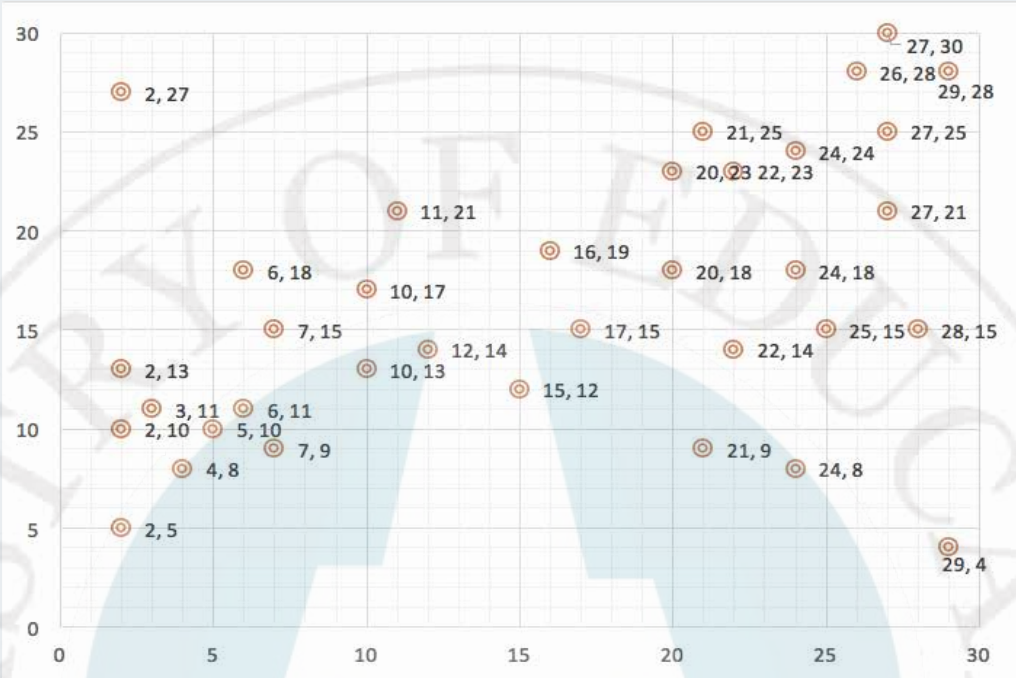
- 1 小叮爸爸預設的兩個地點座標分別為(2,27)和(29,4)
- 2 將這兩個地點連線，找出此連線的中垂線。
(此時所有點被中垂線分隔成兩個區域。)
- 3 分別找出兩區中新的群心A和B，並將其座標填入圖下方空格。
- 4 找出線段 \overline{AB} 的中垂線，此時所有點被中垂線重新分隔成兩個區。
- 5 分別找出兩區的新的群心，並將座標填入圖下方空格。
- 6 重複以上4~5步驟，直到群心座標不再改變即可停止。
(※為了便於計算，我們使用群心的「中垂線」將房子分群，但實際上電腦運作時是藉由計算每間房子與二個群心的距離來分群的哦！)

結果說明：(請依分類結果回答下列問題)

- 1 我一共花了_____回合，才找出最佳的分店地點。
- 2 最後找到的分店地點座標分別為：_____、_____。
- 3 想想看，除了Pizza店要尋找最佳地點之外，還有哪些應用也可以使用這個方法，來尋找最佳的點呢？(請與同學口頭分享)



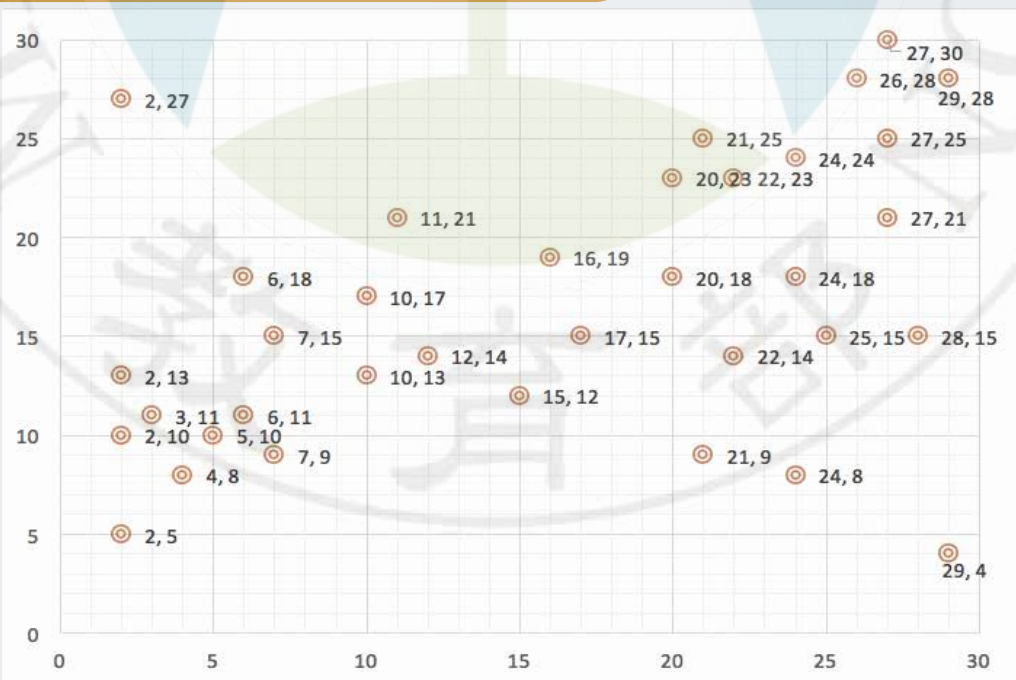
使用K平均分群法找出最佳地點



第 一 回合，新的群心分別是： _____ 、 _____ 。



使用K平均分群法找出最佳地點



第 二 回合，新的群心分別是： _____ 、 _____ 。

第四章：從混沌中搜尋同類-非監督式學習



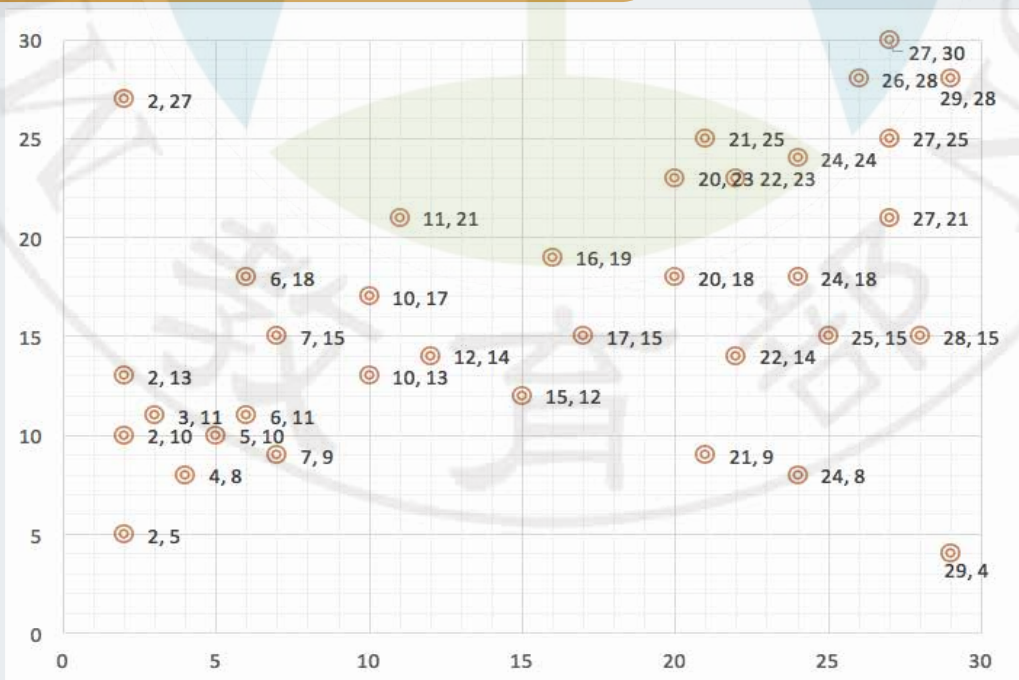
使用K平均分群法找出最佳地點



第 三 回合，新的群心分別是： _____ 、 _____ 。



使用K平均分群法找出最佳地點



第 四 回合，新的群心分別是： _____ 、 _____ 。



「非監督式學習」和「監督式學習」是不同的。「監督式學習」是拿已經標記好的資料來讓電腦做訓練，也可以說監督式學習輸入的資料有標記；而非監督式學習的訓練資料則無任何標記，只能藉由訓練資料的相似性進行分群。例如資料中有大量的貓、狗照片，無須先跟機器說哪些是貓、哪些是狗，由機器自己根據特徵相似性，以分群的方式建構出預測模型。

在非監督式學習中，最具代表性的方法是「K平均分群法」，以下將介紹K平均分群法。

K平均分群法：K平均分群法是使用群聚的方式，將資料分為K個群集，其中K由使用者自行設定，群心是指各群集資料的平均值。

K平均演算法運作原則：

- 1 預先設定欲將資料分成幾個群集（預設K值）。
- 2 從訓練資料中，隨機找出K個資料當作初始群心。
- 3 計算每筆資料到各群心的距離，距離哪個群心最近，該筆資料就被歸類到該群集。
- 4 重新計算群集的平均值，該平均值取代原本的群心，成為新的群心。
- 5 重複執行步驟3、4，直到各群集資料以及各群心不再變動（收斂），分群即結束。

第四章：從混沌中搜尋同類—非監督式學習

EXAMPLE

- 假設有一組整數資料，我們要將其分成三群，資料分別為：
2, 2, 5, 5, 6, 8, 12, 15, 15, 17, 18, 18, 20, 21, 21, 21, 25，在這範圍內我們任取三個群心，分別為：5、15、20。
- 使用K平均分群法分群的流程如下：

第一次分群：

說明 \ 群別	A	B	C
群心	5	15	20
資料分群後	2, 2, 5, 5, 6, 8	12, 15, 15, 17	18, 18, 20, 21, 21, 21, 25
分群後之資料平均 (取四捨五入之值)	5	15	21

(註：為了便於計算，我們四捨五入至整數，但實際上電腦運作是不須要四捨五入的哦！)

第一次分群後之群心分別為：5、15、21

步驟說明：

- 將每筆資料一一判斷距離哪個群心最近，即被分類至該群集。
- 對各群集資料計算平均值，若非整數則將其四捨五入。
(例如：A群資料有2, 2, 5, 5, 6, 8，平均為4.7，四捨五入之後為5)
- 新的平均值，即為下一次分群時新的群心。(第二次分群時群心為5, 15, 21)
- 依此類推，直到群心不再異動為止，分類即結束。

第二次分群：

說明 \ 群別	A	B	C
群心	5	15	21
資料分群後	2, 2, 5, 5, 6, 8	12, 15, 15, 17	18, 18, 20, 21, 21, 21, 25
分群後之資料平均 (取四捨五入之值)	5	15	21

第二次分群後之群心分別為：5、15、21

當執行到第二次分群時，發現各群集與群心不再改變，便是完成分群。分群所需次數會因資料分布及群心選取而異。

K平均分群法也可應用在消費者的購物行為分析。例如：將某購物網站的2000筆顧客資料丟給機器，機器依據顧客資料表之性別、年齡、教育程度、職業、家庭人口組成等，依據各資料之間的相似性將之分群。透過消費行為分析，可以針對消費者可能的需求給予推薦產品，進而刺激消費。我們在瀏覽購物網站時，經常跳出的產品介紹或視窗顯示「你可能會喜歡……」，就是使用這種方法。



書名：和 AI 做朋友-相識篇
副標題：開啓 AI 任意門
著者：陳建中、吉佛慈、蔡淑芬、盧佳君
主編：李建樹
副主編：李育強
執行編輯：陳虹伶、陳瑞翎
封面設計：君傳媒工作室
插圖設計：劉若蕎
發行人：潘文忠
發行：教育部
出版機關：教育部
地址：臺北市中正區中山南路 5 號
電話：02-7736-6666
網址：<https://www.edu.tw/>
出版年月：中華民國 108 年 8 月
版次：初版
其他類型版本說明：本書另有電子版本，請至教育部教育雲網站
(<https://cloud.edu.tw/>) 申請。
定價：新臺幣 120 元

GPN：1010801345

ISBN 978-986-05-9903-9

版權所有，翻印必究

建議由各校彈性課程開課教師自由挑選一到二個單元任務（4 - 8節課）融入而成為跨領域教學活動，或結合各校學生的生活情境問題，成為素養導向的教學活動設計。



AITC

教育部人工智慧技術及應用人才培育計畫
Artificial Intelligence Talent Cultivation Program



教育部AI技術及應用人才培育計畫
中小學分項計畫 8 中小學推廣教育