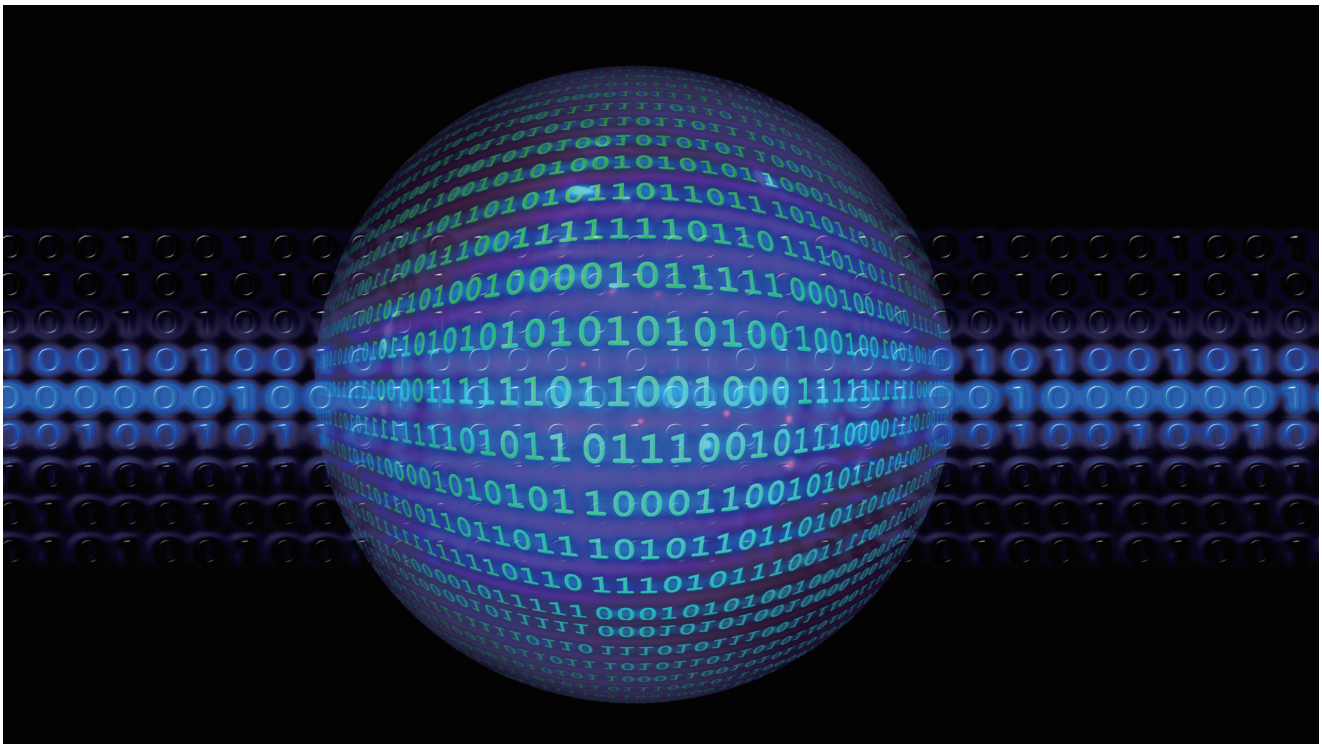
理

什麼是機器學習

●作者：林軒田

機 器 學 習 漫 談

作者簡介：
林軒田現任臺灣大學資訊工程學系教授，也是沛星互動科技的首席資料科學顧問。他的研究領域是機器學習、資料探勘及人工智慧。



(Piqsels)

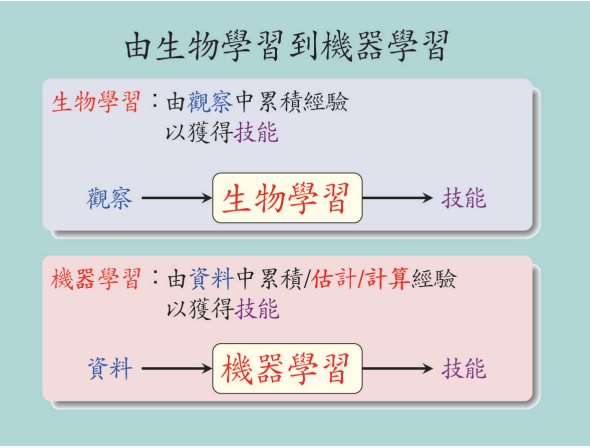
機器學習是個聽起來有點「科幻」的名詞，代表著人類對電腦的一種期望與夢想。在夢想成眞，機器學習成爲熱門顯學的當代，我最常被問到的一種問題，就是「機器學習與 ○○○ 到底有什麼不一樣」。這篇文章就試圖從一些不同的角度出發，告訴大家機器學習是什麼，它與 ○○○ 的關係是什麼，以及有什麼相同或相異之處，希望讓大家有機會用更寬廣更全面的角度看待這門學問。

機器學習與人類學習

要開始介紹什麼是機器學習之前，也許我們可以先想想，人類從小到大是怎麼學習的？人類成長的過程中，最引人注意的學習階段，可能是所謂的嬰兒期的「牙牙學語」。這個學習階段中，嬰兒觀察了身邊的人說話的方式，也觀察了自己可能發出的各種聲響，最終將這些觀察，經過了腦袋的處理，透過模仿、試誤、演繹、歸納等等手段，內化爲自己能掌握的語言體系，從外在看起來，便形成了（母語的）口說技能。

機器學習說穿了，也就是讓機器用它的計算能力去模擬這樣的學習過程。讓機器可以由觀察到的現象出發，透過事先設定好的計算過程，將觀察到的現象內化後，形成適當的技能。這裡觀察

到的現象，在電腦科學上對應的名詞就是資料（data），其中的計算過程就是機器學習的演算法（algorithm），而所得到的技能可以想成是個能實現特定任務的軟體系統（software system），如圖一所示。



圖一

舉例來說，今天若是機器要學會語音辨識這項技能的話，那麼我們可能會收集許多的語音資料給機器，這些語音資料除了數位化的錄音信號之外，可能還包含這些信號的標記資訊，說這些聲音對應到什麼樣的意思。機器有了這些資料後，也許要經過信號處理與分析、空間轉換、目標最佳化等等的演算法，最終得到一個具有語音辨識技能的軟體系統，可以部署在所需的裝置（例如數位音箱）裡頭。

我們怎麼衡量機器學習所得到的軟體系統是否能實現特定任務呢？在機器學習中，我們通常會爲機器定一個效能指標，或更通俗一點的說，就是考試成績。人類學習中，我們常用考試來衡量學生到底學得夠不夠好（雖然考試不見得是讓學生有效學習最好的方式）；在機器學習中，我們也仿效人類學習，定義效能指標來給機器考試。考得好，就是有學到技能；考不好，就是技能還學得不夠。舉例來說，如果我們要用機器學習來做人臉辨識，我們可以用辨識的正確率來當做效能指標，正確率很高，就代表機器學到了人臉辨識的技能；正確率太低，就代表我們還要再努力的教育機器才行。

既然機器學習的技能目標是要得到一個實現特定任務的軟體系統，那透過機器學習方法所得到的軟體系統，和人類自行把一行一行的程式碼寫下來的軟體系統有什麼不一樣呢？人類自行寫下來的軟體系統，是人類學習的結晶，往往需要用許多的規則表現出其中的邏輯思維。但在很多應用中，相對應的思維是「易懂難說」，尤其是許多和人類的常識有關的思維，往往很難表現成邏輯縝密而可以被電腦執行的程式碼；也有一些應用，例如要在極小的時間差之下套利的高頻交易（high-frequency trading），超越了人類能力的物理極限，以致於人類不見得有能力推演出適當的邏輯思維，更遑論要表現成可以被電腦執行的程式碼，在這些不能透過人類學習、人類編寫來實現軟體系統的應用裡，機器學習便可以看成解決是類應用的另類工具。此外，在人類所設計的軟體系統中，這些規則寫在系統裡頭之後，就往往不易變動了；透過機器學習得到的軟體系統，則是可以調變的，給機器學習演算法不同的資料，就可以得到不同的軟體系統，這樣的調變彈性，讓機器學習能夠達成寫一次程式、在不同環境多次使用的應用可能性。

舉例來說，在一個大型的線上歌曲推薦系統體系

（recommendation system）中，要爲各式各樣我們從未謀面的使用者定義出他們都滿意的歌曲推薦規則，是非常困難與耗時的，而且定義出這些規則後，往往也不容易隨著流行趨勢更新。但透過機器學習，演算法可以自動的從資料中爲每個使用者演繹出個人化的推薦清單，並隨著使用者與推薦系統越來越多的互動，洞察出個人喜好與流行趨勢的改變，讓每個使用者的推薦清單都可以與時俱進。

實務上來看，我們常常用三項要素來衡量一個問題是否適用機器學習，或說這個問題使用機器學習是否可行。適用機器學習的第一個可行性要素，是這個問題有沒有可學之處，或者更通俗一點的說，這個問題是否（可能）具有某種規律性，若是一個課題沒有任何規律性，例如要預測一顆全然公平的骰子的結果，不管人或機器，再怎麼學也是白學；當問題有一些潛藏的規律性，機器才有機會由這些規律性學到我們要的技能，例如看似雜亂無章的股票漲跌資料中，可能有些市場交易邏輯的規律性藏在一些蛛絲馬跡之中，而機器如果眞能找出這些規律性，就可以由這些規律性來學到更聰明的高頻交易策略這項技能。

適用機器學習的第二個可行性要素，是這個問題容不容易直接用人類已知的規則或演算法解決。如果有這些規則或解法的話，它們往往比使用機器學習更直接、更簡單、有時還更有效率。例如我們如果想要導航系統指引我們一條由甲地到乙地路程最短的走法，一般電腦科學系的學生都知道我們只需要建立好道路間的連結關係，就能運用有很有效率的最短路徑（shortest path）演算法來算出數學上最短的走法；反之若我們先去收集許多的行車記錄器資料，再讓機器從中去學習，這個由行車記錄器資料到最短路程的解法不但困難，數學上還不見得會給我們最短的解答，反倒變成是「捨近求遠」了。因此能若是一個問題能用人類已知的規則或演算法來得到滿意的解答，我們通常不會傾向於使用機器學習。

適用機器學習的第三個可行性要素，則是我們要有與問題相關的資料。資料是機器學習的原料，是整個機器學習步驟的出發點，有了資料，機器才有辦法開始學習；資料的數量越多、品質越高、內容越豐富、特性越接近我們想要解決的問題，機器越有機會學到解決這個問題的技能；反之若是沒有資料（或是資料太少、資料品質太低），機器是不可能無中生有學到東西的。如果我們想用機器學習解決一個問題，但還沒有資料或資料還不夠好，那麼我們必需先想辦法找出收集資料的策略，才能逐步的讓機器越學越好。舉例來說，在之前所提的推薦系統例子中，若是想做出這個系統的新創公司，也許在草創而毫無資料的時期，必須請使用者先填寫所喜歡的音樂的類型（收集少量的資料），再根據這些類型隨機的推薦一些歌曲給使用者聽並記錄他們的行爲回饋（收集更優質的資料），並透過一些行銷策略累積更多的使用者（收集更大量的資料），這樣才有機會最終實現一個大型的個人化歌曲推薦系統。

舉例來說，我們可以將機器學習運用在貸款核發上，透過申請人所填的各式財務背景資料，來判斷核貸是否對銀行有利可圖還是風險過高。直覺上來看，這個問題的最佳答案的確與這些財務背景資料相關，因此有著可學的規律性，但不見得有一個完

美的公式能直接算出這最佳答案。也就是說，這裡的規律性直覺但未能明文表現，不過銀行裡經過數十年的累積，有著許多與這個規律性有關的歷史資料，這三個要素合起來，顯示機器學習在這個問題上是有著可行性的。

機器學習與人工智慧

人工智慧源自於人類對機器的「期許」，希望具有高度運算能力的機器最終能像人類一樣解決許多困難的問題，甚或超越人類，解決一些人類解決不了的問題。研究機器是否能擁有高度的智慧除了有著哲學上的意義外，工程上也有機會讓電腦成爲更貼心更好用的服務工具。傳統人工智慧，常以機器是否像人與機器是否理性思考兩個主軸，來剖析機器是否具有智慧。這樣的主軸，各自看起來都像是個合理的智慧定義，但將這兩個主軸合起來的時候，隱含表示著聰明的機器要又能夠像人又能夠理性思考。但退一步來看，我們可能會發現，聰明的人類，可能不總是理性思考；理性思考的法匠，搞不好就顯得不近人情不像人了。要又能像人又能理性思考，是人工智慧中常見的邏輯困惑，也大大影響到大眾對人工智慧的接受度。

舉例來說，如果我們想像自駕車要怎樣才叫夠聰明，其中一個邏輯困惑，就是自駕車要如何應對哲學上著名的電車問題。如果一台自駕車開在路上，前方不遠出突然衝出一群天眞無邪的小學生，自駕車經過迅速的計算，發現以目前的車速，緊急煞車或轉向的話，車裡頭的乘客非死即傷；但繼續前進的話，也許乘客在車子的保護機制下，不會受到太嚴重的傷害，但前方國家未來的主人翁們，可能就小命不保了。在這樣的情境下，我們希望自駕車做什麼決定呢？一個比較像人類（例如司機）的決定，是以車內乘客的安全爲第一優先，花了大錢買自駕車的「老闆」們，顯然也會希望自駕車能保護自己珍貴的健康和性命，在這樣的情形下，繼續前進也許是比較像人的選擇；但另一個比較理性的決定，可能是這些小學生不管在人數上、年齡上，都遠比車內少數不再年輕的乘客們有（社會）價值，因此價值衡量的結果，也只好犧牲車裡的乘客了。這個問題告訴我們，要期望自駕車（與許多的人工智慧模組）又能像人，又能理性思考，是難以達成的，不同的應用情境，或也許只是不同的思考角度，都會影響我們覺得需要什麼樣的人工智慧。

也因此我常常覺得當代的人工智慧，重點並不是像不像人，或是否理性。重點是需要是什麼，而這樣的需要能不能被（機器）滿足。也就是說，當代的人工智慧，就是人需智慧。人類依據需求，定義了在特定問題上，想要人工智慧達成的目標，若目標順利達成，人們就覺得機器很聰明，覺得機器很好用。要設定什麼樣的目標，因人而異，也因應用而異。有的應用可能需要更像人一點的目標，例如客服用的對話系統；有的應用可能需要更理性一點的目標，例如自動的股票投資系統；有的應用可能要達到兩者的平衡以取得社會的共識，例如我們提到過的自駕車；又有的應用搞不好不需要像人，但要像個善體人意的伙伴，例如做爲兒童陪伴用途的寵物狗。這些不同的目標，要怎麼達成，也因應用而異，冷氣機裡的智慧溫控功能，可能需要的是好的測溫器，以及適當的控制邏輯，溫度太高就將冷

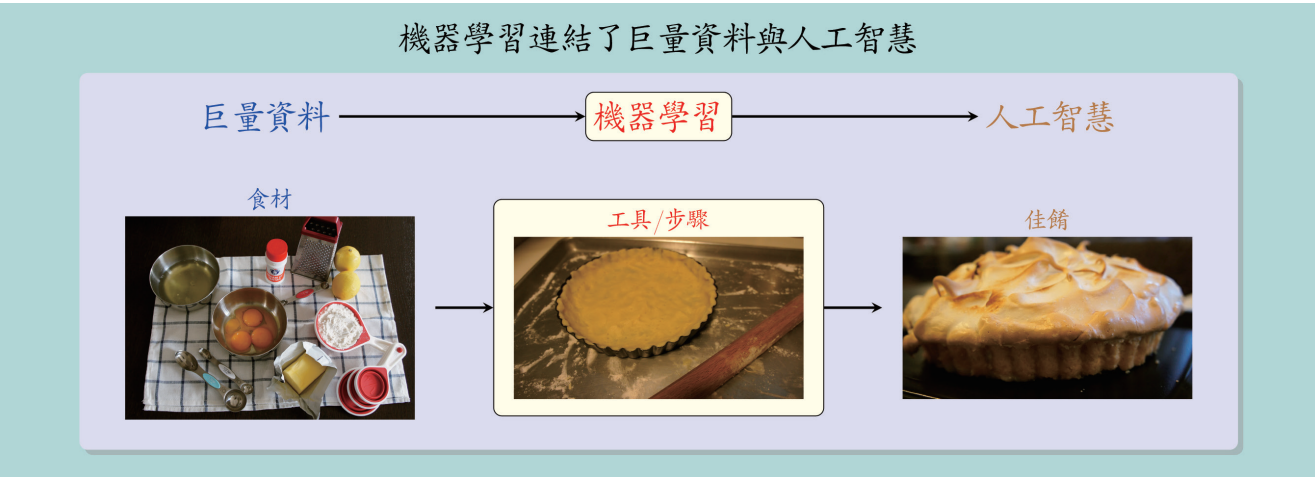
風開得強一些，溫度太低就就冷風關得小一些；自動的人臉門禁系統，可能就需要複雜的物件切割、辨識、記錄與搜尋的能力。

人工智慧是個發展了非常久的領域，裡頭有非常多經典的哲學與方法來讓電腦有智慧，機器學習是其中非常重要的一派方法，在當代人工智慧中甚至可以說是最主流的一派方法。舉例來說，在解決各式棋類問題上，傳統人工智慧研究者常常想辦法依人類知識定義棋盤每個盤面的分數後，再利用樹狀的展開（有點類似人類在腦袋中推演盤面的發展），來搜尋出最佳的下法；而當代人工智慧，例如知名的 AlphaGo 圍棋程式，則可以透過大量的生成各式的對奕資料，用機器學習自動估算分數並決定最佳的下法，並隨著自我練習的生成過程持續進化。在許多問題上，機器學習已顯現出它在當代人工智慧中的強大競爭力。

我常用大廚上菜的比喻來讓大家了解機器學習與當代人工智慧的關係，如圖二所示。人工智慧像是廚師想端出來給大家的一道道佳餚，要完成這道佳餚，不可或缺的部份，就是做爲原料的食材，對應到的就是資料。而食材與佳餚之間的工具與步驟，所謂的烹飪，就是廣義的機器學習，而執行這些廚藝的特級廚師小當家，自然就像是我這樣的所謂資料科學家或機器學習科學家了。

更進一步的來說，我們先前說到機器學習是由資料出發，透過機器學習的演算法，而形成實現特定任務的軟體系統。這樣的軟體系統若是能有效提升與人工智慧有關的效能指標，我們便達成了人工智慧。也就是說，機器學習所產出的軟體系統，即是人工智慧的一種實現。隨著資料產生、儲存和處理的成本越來越低，以機器學習來實現人工智慧的可行性也就越來越高，也因此在當代人工智慧中，機器學習成爲最主流的一派方法。

當然，以上的比喻只把資料與人工智慧中間的橋樑統稱爲機器學習，其實是過度簡化了。狹義的機器學習，可能只對應到烹飪裡頭熱火烹調的那部份



圖二。（圖中照片 Flickr，Andrea Goh）

步驟，但眞正要實現一個好的人工智慧系統，要做出一道好菜，也許得從資料的整理（就像是洗菜與食材的處理）出發，而這些資料整理的工夫，雖然對於機器學習的好壞無比的重要，卻通常不被看成核心的機器學習步驟；又或者煮出美味的食物後，一道好菜還得靠著適當的盛裝與擺盤（就像是軟體系統的佈署）才能眞正讓顧客得到佳餚的享受；又或者有些佳餚，可以直接從食材處理完後，跳過熱火烹調，用其他的技術處理後，就直接擺盤上桌，這也就是說，機器學習不見得是由資料實現人工智慧的唯一選擇。機器學習固然在當代的人工智慧系統中扮演不可或缺的角色，但要實現一個成功的人工智慧系統，絕不能只有機器學習，而需要通盤納入所有可能的工具在系統當中。

機器學習與資料探勘

記得我們提到機器學習的出發點是資料，而目標是要讓機器透過資料學到某種技能，或者概略的說就是學到一個人工智慧系統。同樣用資料出發的一個領域叫做資料探勘，它的目標是要讓機器自動或半自動的爲人類找出資料裡頭「有趣」之處。舉例來說，資料裡頭有沒有哪些部份具有很強的規律性、關聯性或群聚性？有沒有哪些資料的值有異常值得懷疑之處？如何視覺化的表現資料之間的關係？

舉例來說，一個很經典的資料探勘問題，是要找出眾多超市顧客中，最常「一起購買」的商品是哪些，這個問題經由一個廣爲流傳的都市傳說尿布與啤酒的關聯性渲染之後，成爲資料探勘的經典代表。尿布與啤酒的傳說中，一家超市從銷售記錄中發現，許多顧客星期五下班前會去超市帶一些啤酒回家看球賽，買啤酒的同時，他們也會順手替家中的嬰兒買一些尿布，於是超市運用了這個資料探勘的結果，設計爲這類顧客所設計的貨架擺放方式，以提昇顧客的購物體驗並刺激銷售。經過考證，這個傳說似乎並非眞人眞事，但資料探勘在現實的商業問題中，的確有助於發現一些資料中有趣的關聯性，進而協助人類改進商業流程或行銷策略。

也就是說，同樣從資料出發，機器學習與資料探勘，爲想要的輸出設定了不同的目標，前者的目標是有用的技能，後者的目標是有趣的發現。在一些應用中，有用的技能和有趣的發現，其實是一體的兩面，而這時機器學習和資料探勘便只是一體兩面，沒什麼不同。舉例來說，如果今天給定台灣股市最近二十年的資料，機器學習的研究者可能會想要機器學到預測股市明天會漲會跌的技能，而資勘的研究者可能會想要機器找出股市會漲會跌的模式，要叫技能還是叫模式，其實只是換句話說而已，這時機器學習和資料探勘想做的事情是一致的。

此外，想由資料中學到技能（機器學習），常常需要先觀察資料中的有趣之處（資料探勘）；而若從資料中學到一些技能（例如：預測股市漲跌）之後，對這些技能更進一步的分析（例如：看看所學到的技能在哪些情形會預測股市上漲）可能也可以告訴我們更多資料中的有趣之處。因此實務上來看，機器學習和資料探勘有非常高的關聯性，沒有辦法（也不需要）劃出一條明確的界線來做區分。

傳統的資料探勘，除了探索資料中有趣之處外，

因爲所處理的資料量往往十分龐大，還會特別著重這些資料在儲存、取用與查詢上的效率，這些面相發展爲資料庫設計這門非常重要的學問，而這門學問在傳統的機器學習中比較少著墨。然而隨著所謂巨量資料的發展，當代機器學習也有許多的方法，需要配合適當的資料庫設計與相對應的資料探勘技巧，才能迅速的整合各式異質性的資料，來學到更強大的技能，也讓機器學習和資料探勘的關係，變得更加密不可分。

機器學習與應用數學

接下來，我們來談談機器學習與應用數學。先前說到機器學習是由資料出發的學問，而在應用數學中，同樣也是由資料出發的學問，就是統計學。統計學希望在某種模型的假設下，由資料中估計資料分佈中的某項特性。舉例來說，由一百次獨立的電話調查結果，估計每個候選人的支持度。而機器學習，則希望由運用資料中的潛藏規律性，得到實現特定任務的一段程式碼（技能），或說一個函數。如果我們把機器學習想得到的函數，和統計學中想估計的特性，看成同一件事，而把機器學習中的潛藏規律性與統計學中的模型假設也看成同一件事，那麼機器學習跟統計學也就完全一致了，或者很粗略的說，有很大一部份的機器學習可以看成（推論）統計學中往電腦科學延伸的一門學問。

不過機器學習與統計學重視的面向還是有一些不一樣，其中一項不同之處，就是統計學的出發點是數學，所以會更注重在什麼樣的狀況下，能透過適當的數學假設，而得到能被數學證明的結果；而機器學習的出發點是電腦科學，所以會更重視怎樣能有效的算出我們想要的函數，也因此許多當代機器學習的發展會更偏向方法設計與應用，而統計學中相對應的內容會更偏向方法分析與理論。注意到這樣的偏向並不是個黑白分明的區分，兩個領域持續的還在互相合作與互相刺激，許多經典的統計結果與模型（例如著名的羅吉斯迴歸〔logistic regression〕）已大量的被引入機器學習使用，許多新興機器學習的需求（例如集成學習〔ensemble learning〕）也吸引了許多優秀的統計學家投入並產生了新穎的結果（例如隨機森林〔random forest〕）。

在機器學習裡，方才提到會更重視怎樣能有效的算出我們想要的函數，而這個算出的過程，通常對應到對某個效能指標（如前所述）做最佳化。因此另一個與機器學習高度相關的應用數學領域，就是數值最佳化。許多機器學習的問題只要能根據效能指標寫成某個需要最小化的目標函數，那接下來的典型步驟便是簡單的算出這個目標函數的梯度後，使用梯度下降法（gradient descent）將目標函數最佳化。數值最佳化這個應用數學的領域，數十年的研究成果，讓這個步驟有著許多不同的變化。舉例來說，當目標函數搭配有限制的條件的時候，也許需要使用投射梯度（projected gradient）而非一般梯度；當我們有機會能計算目標函數的二階導數（而非僅是梯度這個一階導數）的時候，也許可以使用牛頓法（Newton method）來求解；當目標函數和限制的條件有著特別的型式的時候，也許可以轉成數值最佳化中經典的線性規劃（linear programming）、二次規劃（quadratic programming）或半正定規劃

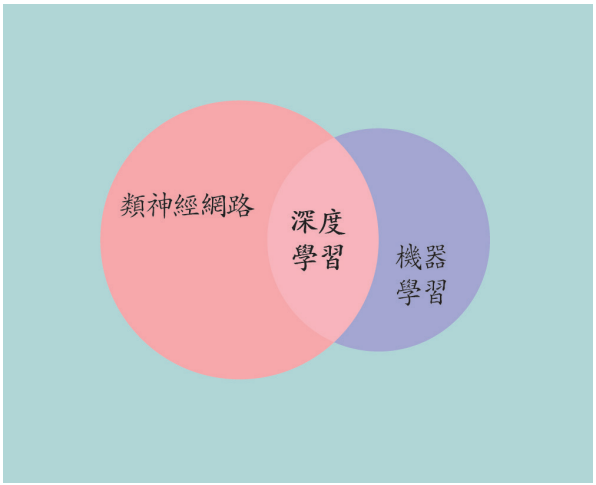
（semi-definite programming）來求解。

機器學習對應到的數值最佳化問題，通常有一些不同於一般最佳化問題的特性。舉例來說，許多的目標函數，都可以表示成每筆資料所對應的效能指標的疊加，這時我們便可以從疊加中隨機取樣出數筆資料，來計算出它們的效能指標上的隨機梯度（stochastic gradient），並在梯度下降法中以隨機梯度取代傳統梯度；隨機梯度雖然比傳統梯度來得不穩定，但當機器學習運用在巨量資料上時，傳統梯度在計算上會碰到很大的瓶頸，而隨機梯度成爲實務上十分便利的計算捷徑，這樣的便利性，讓隨機梯度下降已成爲深度學習中不可或缺的一項工具。像隨機梯度下降這樣結合機器學習特性的數值最佳化方法，是當代機器學習演進過程中，非常重要的一項推力。

最後，我們用一個經典的機器學習方法，爲大家整合機器學習與應用數學的關係，支撐向量機（support vector machine）是個經典的機器學習方法，它源於萬普尼克（Vladimir Vapnik）和澤范蘭傑斯（Alexey Chervonenkis）兩位統計學家對於由統計出發的機器學習理論的探討[1]，並依所推導出的一些理論結果設計出一個新穎的機器學習模型。這個模型對應到一個特殊的二次規劃問題，可以運用二次規劃問題的對偶性質，用應用數學中各種已知的二次規劃的工具來對原問題或其對偶問題求解，而也可以進一步利用這個問題中特殊的性質，例如對偶問題解答的稀疏性，設計特殊的最佳化演算法，讓機器能有效的運用更大量的資料來學習。

機器學習與深度學習

最後，我們來談談深度學習，它是近年來非常熱門的名詞，源自於經典的類神經網路領域，這個領域結合了電腦科學及神經科學，試圖建立一些模型，來模仿或模擬生物體中用一堆神經元互相連結而產生的計算、思考、記憶、學習等現象。其中有關學習的部份，經過數十年來的研究，加入了許多工程的設計，已與生物體中所觀察到的作法不盡相同，但卻演進成機器學習中一支非常重要的模型家族，也就是我們今天所稱的深度學習，如圖三所示。



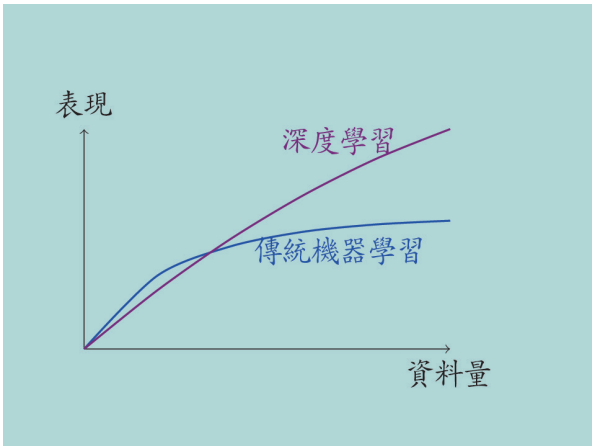
圖三。

從這樣的角度來看，深度學習其實就只是位在機器學習與類神經網路交集處的一支模型家族而已，那麼跟機器學習領域中其他的模型比起來，這個家

[1] [編註] VC 理論（Vapnik-Chervonenkis theory）。

族有什麼特別之處呢？首先，從應用端來看，深度學習的模型在一些相當困難的問題上，一舉突破了人類幾十年來都無法攻克的一些效能障礙。舉例來說，在 2012 年時，一個深度學習中的卷積神經網路（Convolutional Neural Network）模型，在一個 1000 個類別的影像分類比賽上，達到了猜五次 15.3% 的錯誤率，得到了第一名，這在當年是非常優異的成績（與當年第二名的差距超過 10%）；在 2016 年時，Google DeepMind 的 AlphaGo 程式，使用了卷積神經網路以及自我對弈等訓練技巧，以五戰四勝擊敗了人類的頂尖圍棋棋士李世乭，達到大幅度壓制人類的人工智慧表現。這些表現顯示出深度學習具有非常豐富的潛力，也讓許多學界與業界研究者，有更強的信心想要使用這個模型家族來解決更多困難的問題。

而從技術端來看，深度學習的深，指的是這個模型是像樂高積木一樣，使用簡單的元件一層一層組合起來的，組合了越多層，我們就稱之爲越深。這裡的元件，模擬了生物體裡頭的一個小小神經元，單一的神經元只表現了一個簡單的函數，但當我們將每個神經元層層組合之後，便能轉化爲更複雜的函數，圖三提供。這樣的描述看似簡單，但其實有很多技術端的挑戰，但如當有這麼多神經元，每個神經元又各自帶有許多待學習的參數的時候，深度學習的問題，便可以看成一個超高維度、非線性、且非凸（non-convex）的最佳化問題，這在最佳化領域中，是個極不好解的問題；而這個超高維度的的問題，在用來學習的資料量過少的時候，可能會讓電腦有機會把資料的雜訊通通都一起背起來了，產生了統計上所謂的過適（overfitting）現象，這需要靠更巨量的資料，或設計出更強健的深度學習模型來解決；但若要運用巨量的資料在超高維度的問題上，計算上還是有可能會碰到瓶頸，於是需要去思考如何運用軟體平台（例如多執行緒運算）或硬體平台（例如適合平行運算的圖形加速器）加速所需的計算。如果把深度學習比喻爲機器學習中的一把火的話，這些技術上的突破，就像是讓我們能夠點起了火、安全的不燒到自己、並進一步的將火用在有用的地方（例如汽車的引擎）。當我們能順利點起這把火，深度學習就可以運用大量資料與大量運算，持續提昇效能表現，這是傳統機器學習所不能及的，如圖四所示。



圖四。

類神經網路的研究，在幾十年前就已開始，但當年難以使用具有多層的神經元的深度類神經網路，因爲當時沒有夠多的資料、沒有夠快的運算能力、也沒有夠搭配的學習方法。這樣的限制，讓類神經網路有一陣子陷入無人聞問的寒冬。這個領域在當

代隨著資料量、運算能力及學習方法設計一同進展之後，一些研究者爲了擺脫先前類神經網路這個名詞給人的負面印象，就重新賦予這個模型家族深度學習之名。

最後，我們可以很快的透過深度學習，來理解本文中跟大家解釋的各種面相。深度學習是機器學習中的一個模型家族；它起源於模仿人類學習，同樣試圖由資料透過演算法的內化來學到某項技能；它強大的模型潛力，讓它能在許多困難的人工智慧問題上，成爲主流的方法；深度學習的潛力來自於資料中找出從簡單到複雜的規律性，它和資料探勘在巨量資料上要做的事情其實是一體兩面，相輔相成；在建立模型和找出規律性的過程中，需要連結到應用數學中統計學、數值最佳化等方向，並於電腦科學中在各式軟硬體上的高效能程式設計相結合。

延伸閱讀

▶阿布 - 穆斯塔法（Yaser Abu-Mustafa）、馬頓 - 伊斯梅爾（Malik Magdon-Ismail）和本文作者合著的《從資料中學習：短期課程》（*Learning from Data: A Short Course*, 2012, AMLBook）。這本書是曾多次名列亞馬遜暢銷排行榜第一名的機器學習領域的入門教科書，讀者可以通過閱讀本書來認識機器學習領域的所有基礎知識。也可以同時配合
https://www.csie.ntu.edu.tw/~htlin/mooc/
的鏈結觀看作者在臺大磨課師（MOOC）所開設的「機器學習基石」與「機器學習技術」課程。

▶本文作者在臺大科學教育發展中心 CASE 的公開科普演講，探索 20-2 講座〈開發機器的學習潛能 —— 鑽牛角尖或舉一反三？〉：
https://www.youtube.com/watch?v=whuqqheSMMI

▶本刊曾於第 10 期與第 17 期分別規劃了「人工智慧與 AlphaGo」「機器學習」的專題，讀者可以參照閱讀。