歡迎學習“模組 2：機器學習簡介”。

本模組將介紹以下主題：

- 什麼是機器學習？
- 利用機器學習解決業務問題
- 機器學習術語和過程
- 機器學習工具概覽
- 機器學習挑戰

您需要完成一個知識測驗，以測試您對本模組中的關鍵概念的理解程度。

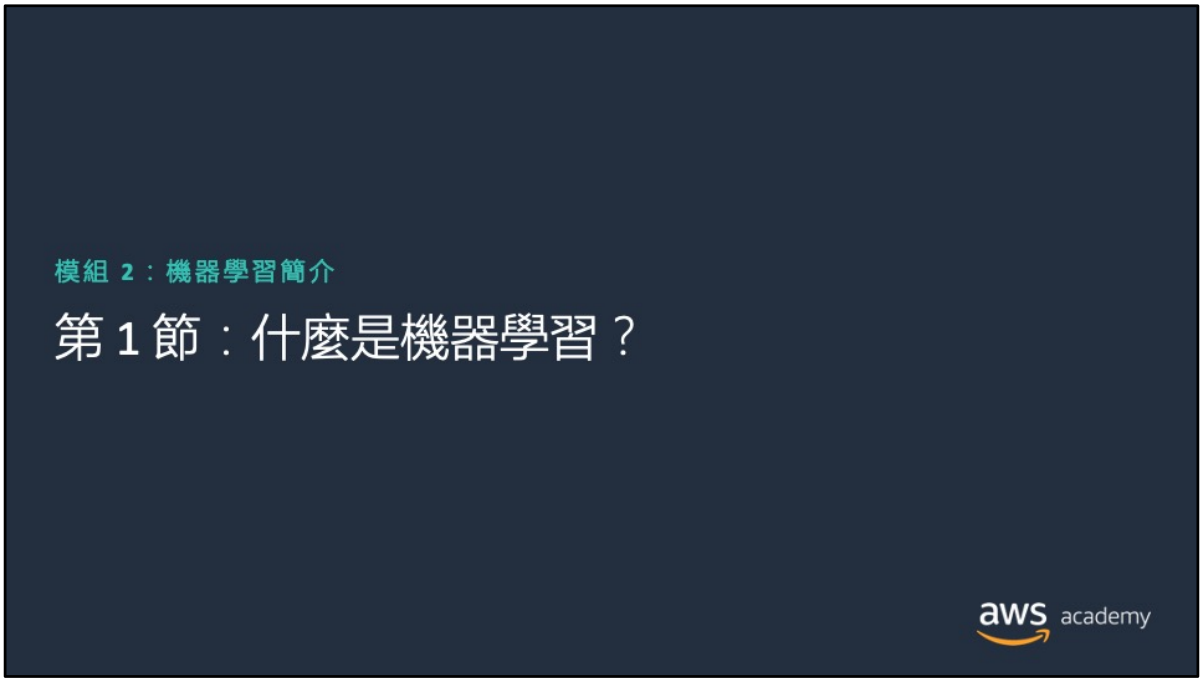
模組目標

學完本模組後，您應該能夠：

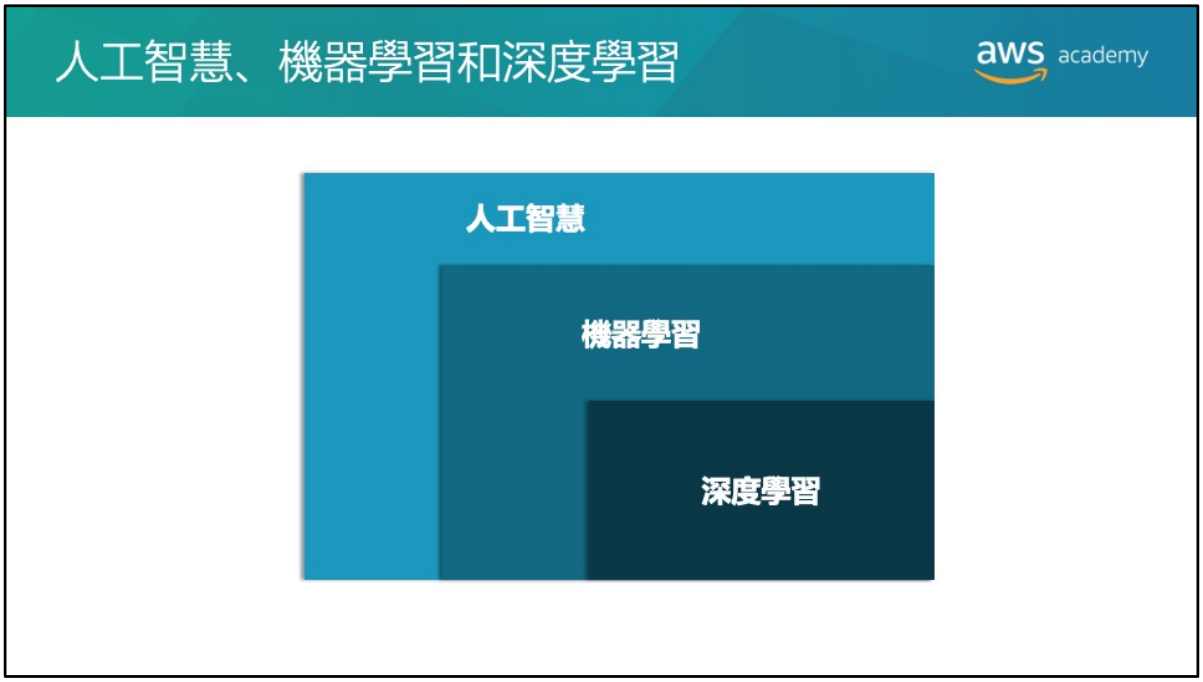
- 認識到機器學習和深度學習是人工智慧的一部分
- 解釋人工智慧和機器學習術語
- 確定如何使用機器學習解決業務問題
- 說明機器學習的過程
- 列出資料科學家可用的工具
- 確定何時使用機器學習，而非傳統的軟體發展方法

完成本模組後，您應能夠：

- 認識到機器學習和深度學習是人工智慧的一部分
- 解釋人工智慧和機器學習術語
- 確定如何使用機器學習解決業務問題
- 說明機器學習的過程
- 列出資料科學家可用的工具
- 確定何時使用機器學習，而非傳統的軟體發展方法

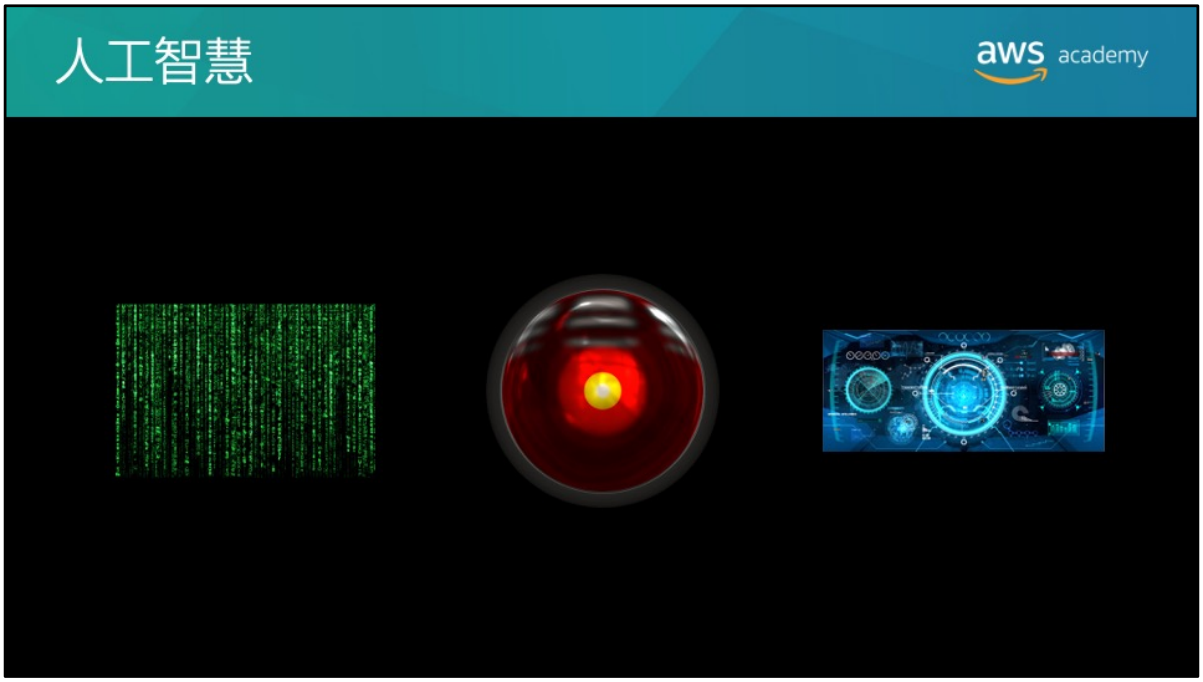


介紹“第 1 節：什麼是機器學習？”



本課程介紹了機器學習 (ML)，並且展示了機器學習在規模更大的人工智慧 (AI) 領域中的定位。

機器學習是 AI 的一個子集，AI 是電腦科學的廣義分支，用於構建可以執行人工任務的機器。深度學習本身是機器學習的一個子領域。為了理解這些概念在何處交匯，您將學習每個領域。



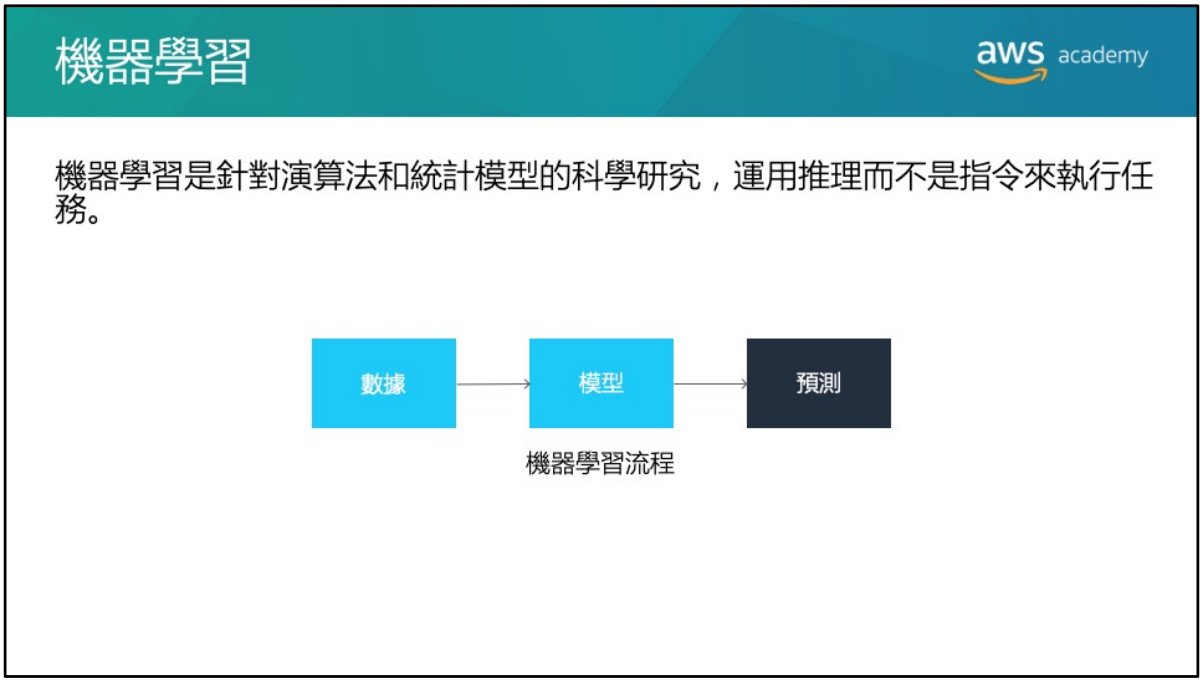
機器學習是稱為人工智慧 (AI) 的更廣義的電腦科學領域的子集。

AI 關乎於製造能夠執行通常由人類執行的任務的機器。在現代文明中，AI 常常亮相在電影或小說中。您可能會想起某些科幻電影或電視節目中的 AI，它們可以控制未來世界，能夠自行智慧地行動 - 有些時候會對社會或周圍的人產生負面影響。這些 AI 最初是作為感知環境的電腦代理，並執行某些行動來實現特定目標。但是，其中一些虛構的 AI 的行動方式並非其創建者最初設想的結果。

其他虛構的 AI 要較為善良或正面：它們能更好地與人類合作，不過它們的用途也要更加廣泛。這些類型的通用 AI 就屬於通用人工智慧 (AGI) 範疇。它們有能力學習或理解人類可以理解的任何任務。

AI 問題通常涉及許多研究領域：自然語言處理、推理、知識表示、學習、感知以及與物理環境交互。對於生活在現實世界中的人來說，AI 還沒有成為現實。但是，每一年，在每一個領域中，AI 都越來越接近現實。

您可能也讀到或看到過對於創建 AI 倫理的評論。並非所有觀點都是正面的 – 也許部分是因為人們擔心虛構故事中的那些會摧毀人類或者將人類用作動力來源的 AI。有些人可能還會考慮大規模失業的風險，畢竟智慧型機器可以不用休息、連續工作。



您可能會發現機器學習的許多定義。不存在標準的定義，下面我們來介紹幾種機器學習的定義。

*機器學習*是針對演算法和統計模型的科學研究，運用推理而不是指令來執行任務。為了幫助您更好地理解這一點，請看下面這個具體示例。

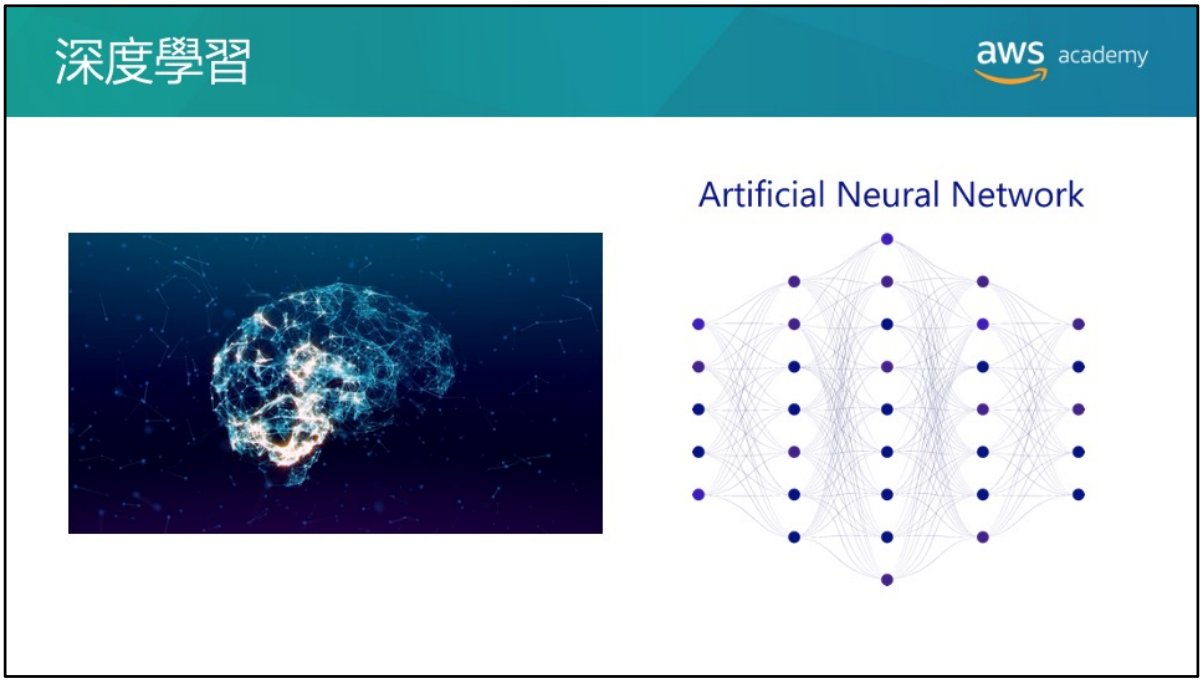
假設您必須編寫一個應用程式，來確定電子郵件是否為垃圾郵件。如果沒有機器學習，您只能編寫一系列複雜的決策語句（考慮 *if/else* 語句）。或許您會使用主題或正文中的單詞、連結數量以及電子郵件的長度來確定電子郵件是否為垃圾郵件。編纂如此廣泛的規則以涵蓋所有可能性將是十分困難而費力的。但借助機器學習，您可以使用已經標記為*垃圾郵件*或*非垃圾郵件*的電子郵寄清單來訓練機器學習模型。該模型將瞭解哪些單詞、長度和其他指標模式可以很好地預測垃圾郵件。在您向模型展示其以前從未見過的電子郵件時，模型會預測它屬於*垃圾郵件*還是*非垃圾郵件*。

機器學習領域的先鋒 Tom Mitchell 給出了這樣的定義：“如果電腦對某類任務 (T) 中任務的執行績效（按績效指標 (P) 衡量）隨經驗 (E) 的增加而有所提高，那麼就可以說，電腦程式可以從經驗 (E) 中學習有關這類任務 (T) 和績效指標 (P) 的資訊。”（Mitchell, Tom.1997 年。*機器學習。*McGraw Hill。第 2 頁。）

如果您對垃圾郵件應用此概念，則字母 *E*、*T* 和 *P* 分別表示如下含義：

- E – 指示是否為垃圾郵件的電子郵件
- T – 識別垃圾郵件的任務
- P – 先前未見過的電子郵件屬於垃圾郵件的可能性。

如果您覺得這種思路難以理解，請不要擔心。本模組稍後部分更完整地介紹了這些步驟。



深度學習代表了 AI 和 ML 功能的重大飛躍。深度學習背後的原理源於人類大腦的工作原理。儘管實現方式不同，但*人工神經網路 (ANN)* 的靈感來自於大腦中的生物神經元。

人工神經元有一個或多個輸入和一個輸出。這些神經元會根據輸入的變換觸發（或啟動其輸出）。

神經網路由這些人工神經元的層組成，各層之間具有連接。通常，網路具有輸入層、輸出層和隱藏層。

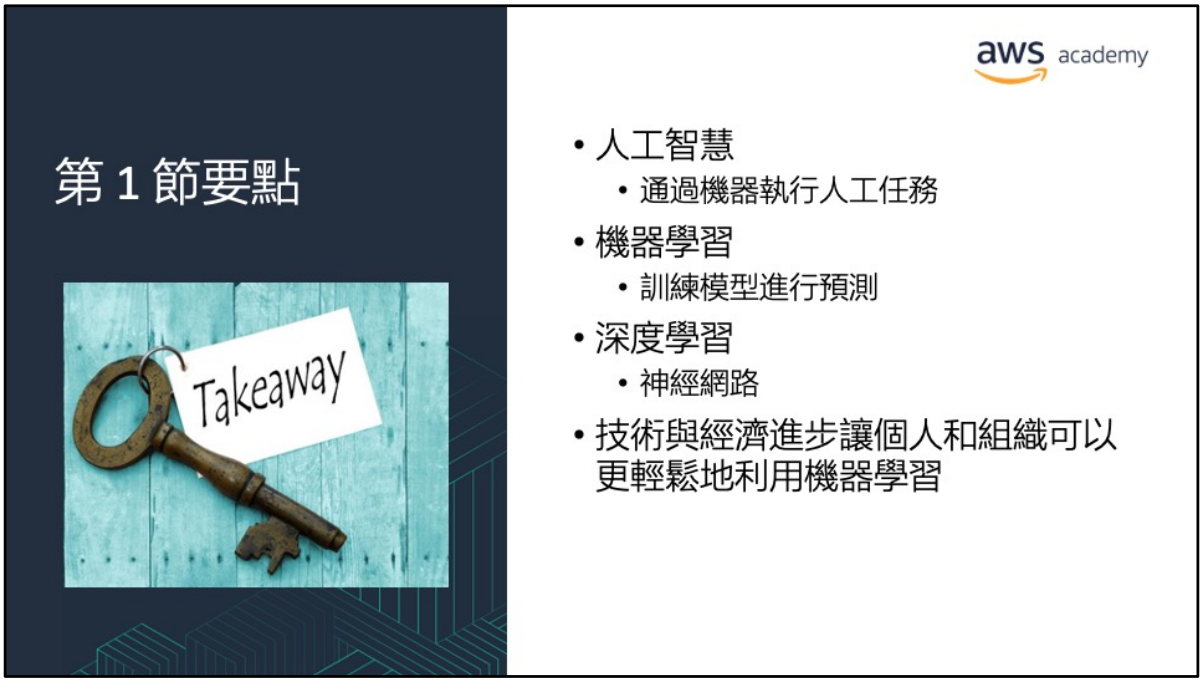
一個神經元的輸出連接著下一層中所有神經元的輸入。然後，要求網路解決問題。輸入層會基於訓練資料填充。神經元會在整個層中啟動，直到在輸出層中給出答案為止。然後，測量輸出結果的精確度。如果輸出未達到您的閾值，則重複執行訓練，但神經元之間的連接權重會稍有變化。它將持續重複此過程。每一次，它都會加強導致成功的連接，並減少削弱失敗的連接。

正如您將在本課程中看到的那樣，機器學習從業者要花費大量時間來調優 ML 模型。他們選擇最佳的資料特徵進行訓練，並選擇結果最佳的模型。深度學習從業者幾乎不會在這些任務上耗費什麼時間。他們的時間花費在使用不同的 ANN 架構對資料進行建模上。

雖然深度學習的理論可以追溯到幾十年前，但運行深度學習問題所必需的硬體直到最近才普及使用。現在，有了這些硬體，您就可以使用深度學習來解決比以往更複雜的問題。

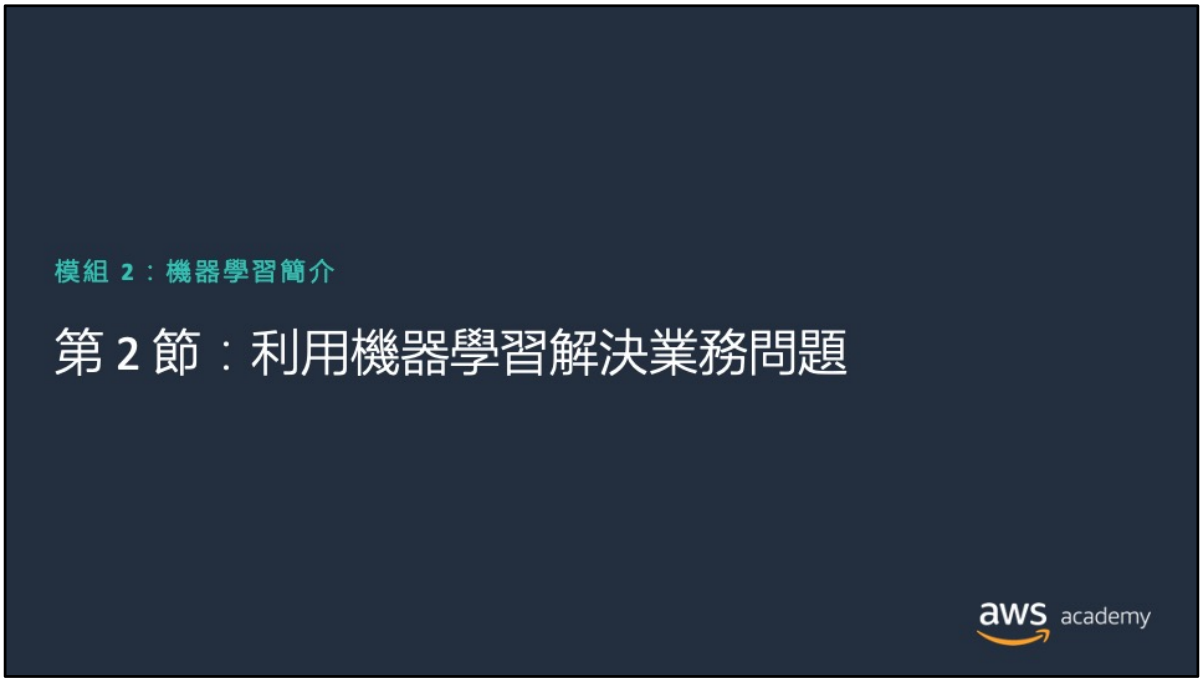
機器學習是近年來才走入主流的。2000 年代中期，機器學習和深度學習開始迅猛發展，部分原因在於摩爾定律和雲計算的興起。這讓人們可以更輕鬆地獲得容量更大、速度更快、價格更便宜的計算和存儲能力。過去，您需要斥鉅資購買和運行大型計算集群，而現在只需租用幾個小時的計算能力，費用遠遠低於過去。

2012 年，以圖像識別為主題的機器學習競賽 ImageNet Large Scale Visual Recognition Challenge 揭開了神經網路投入實際應用的序幕。其準確率一舉躍升至 82%，此後一直穩步上升。2015 年，神經網路的表現就已經超過了人類。



本模組中這部分內容的要點包括：

- 人工智慧的領域範圍要更加廣闊，不僅限於製造機器，更延伸到了執行人工任務。
- 機器學習是 AI 的一個子集。它著重於使用資料來訓練 ML 模型，以便模型能夠做出預測。
- 深度學習是一種靈感源自人類生物學的技術。它使用神經元層來構建網路，解決問題。
- 技術、雲計算和演算法開發的進步讓機器學習功能和應用日漸增加。



介紹“第 2 節：利用機器學習解決業務問題”。

常見業務使用案例

垃圾郵件與常規電子郵件

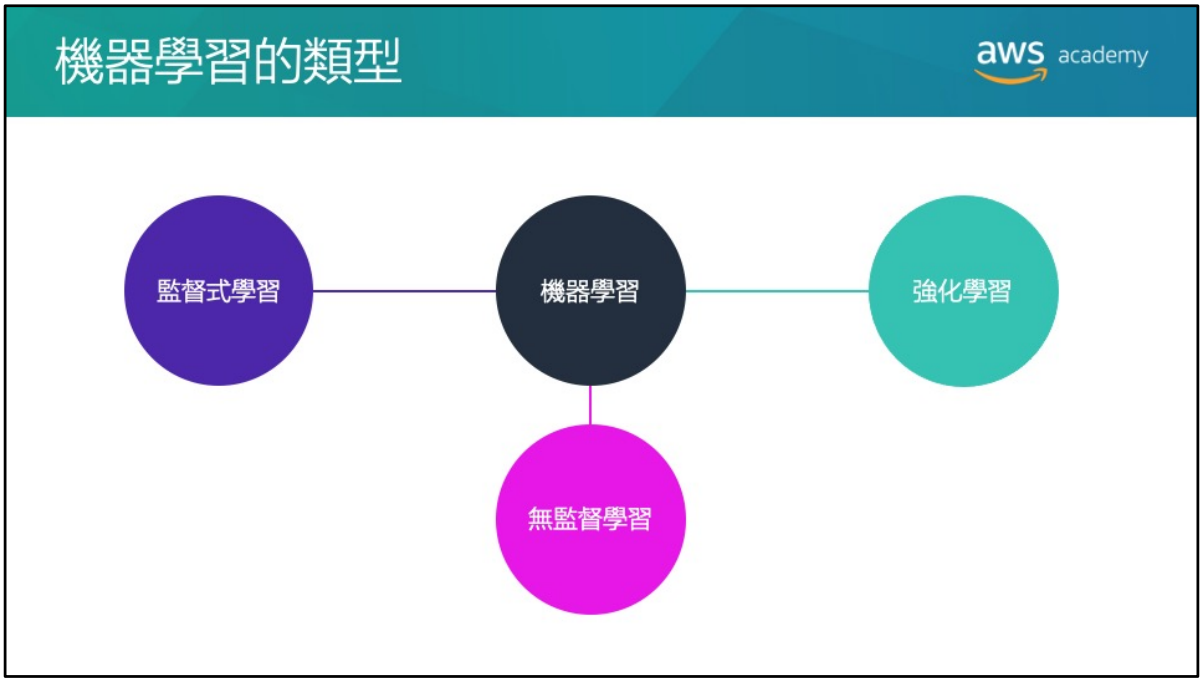
推薦

欺詐

在每個人的數位化生活中，都會用到機器學習。這裡給出了幾個示例：

- 垃圾郵件 – 您的垃圾郵件篩檢程式就是 ML 程式的結果，這種 ML 程式使用示例垃圾郵件和常規電子郵件進行了訓練。
- 推薦 – ML 程式根據您閱讀過的圖書或購買過的商品，預測您可能需要的其他圖書或商品。類似地，這類 ML 程式使用讀者習慣和消費者購物歷史資料進行過訓練。
- 信用卡欺詐 – 同樣，ML 程式也接受過有關示例欺詐交易和合法交易的訓練。

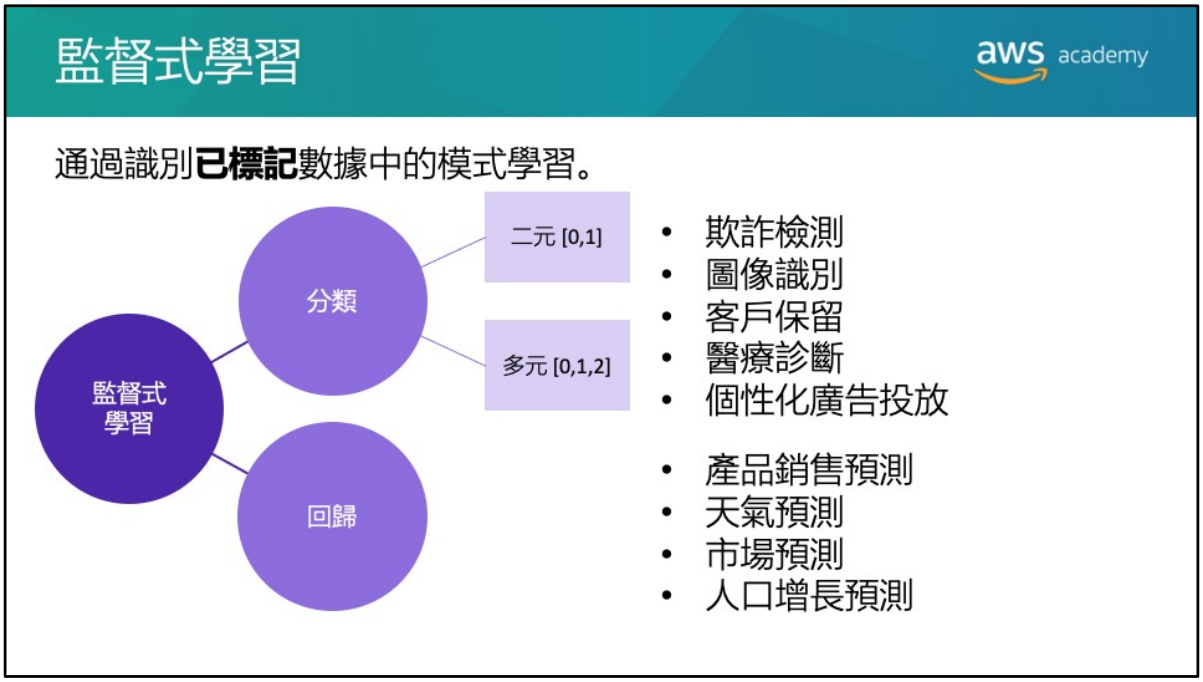
還有更多相關示例不勝枚舉，包括社交媒體應用程式中使用人臉檢測對照片進行分組，在腦部掃描中檢測出腦部腫瘤或在 X 光片中發現異常。



機器學習分為三個主要類型。

第一種類型是*監督式學習*，即模型使用已知輸入和輸出來泛化未來輸出。第二種類型是*無監督學習*，即模型不知道輸入和輸出，在沒有說明的情況下從資料中發現模式。第三種類型*強化學習*，即模型與其環境交互，學習採取能夠獲得最大回報的行動。

瞭解不同的 ML 類型非常重要，因為這可以指導您選擇對解決業務問題有意義的演算法。



*監督式學習*是一種常見的 ML，因為它的應用非常廣泛。這之所以稱為監督式學習，是因為需要監督員（就像是老師）來為模型展示正確的答案。

像學員一樣，監督演算法需要通過示例學習。實際上，它需要一名“老師”來使用訓練資料說明它確定輸入和輸出之間的模式和關係。“這是一張汽車圖片。這是另一張汽車圖片。”模型使用這些打標資料進行訓練，從而準確識別此前從未見過的圖片中的汽車。

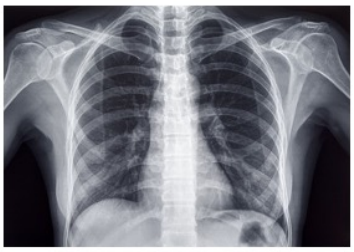
但是，監督式學習中有不同類型的問題。這些問題可大致分為兩類：分類和回歸。

*分類*問題有兩種類型。第一種類型是*二元分類問題*。回想一下前面關於識別欺詐交易的示例。本示例中的目標變數只有兩個選項：*欺詐*或*非欺詐*。這是一個二元分類問題示例，因為您要將觀測值分為兩個類別。

此外還有多類別分類問題。這種 ML 問題將觀測值分為三個或更多類別。假設您有一個 ML 模型，該模型可以預測客戶致電聯繫您的店鋪的原因。其目的是減少將客戶轉給正確的客戶支援部門之前所需的轉接次數。在本示例中，不同的客戶支援部門代表不同的潛在目標變數 – 可能是許多不同的部門。

此外還有*回歸*問題。在回歸問題中，您不再將輸入映射到指定數量的類別。相反，您要將輸入值映射到整數這樣的持續值。讓我們來看一個預測公司股價的 ML 回歸問題示例。例如，一個基於回歸的演算法將預測到明天公司的股價可能會從每股 113 USD 上升到每股 127 USD。

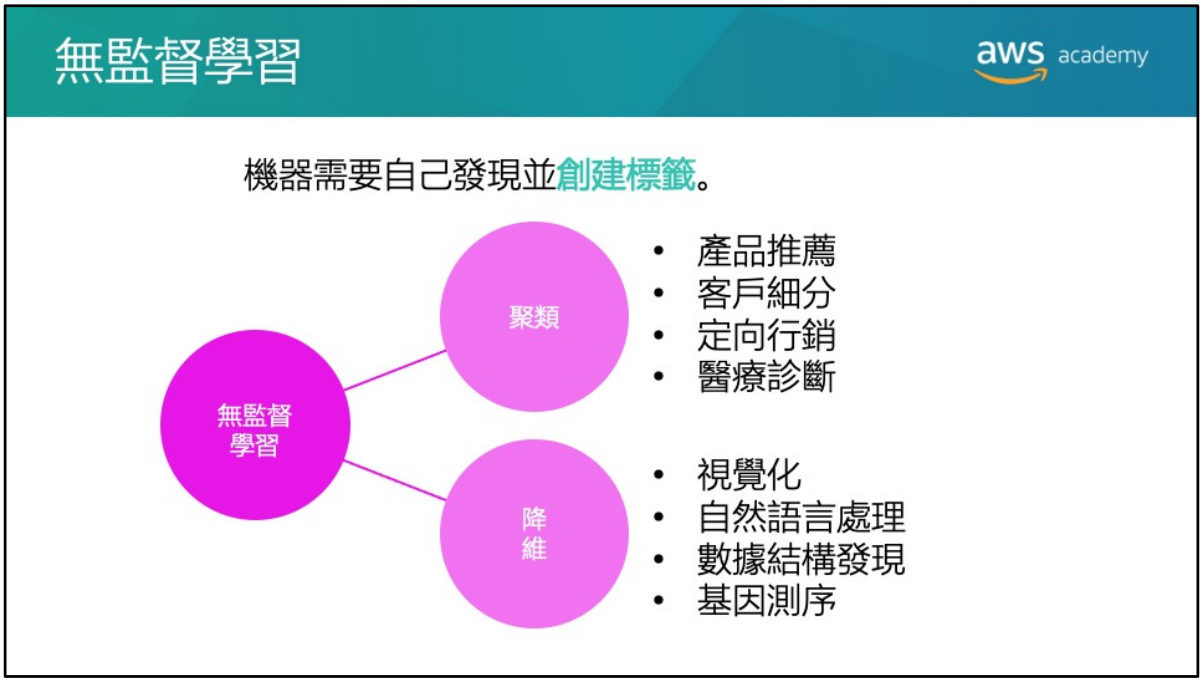
電腦視覺



電腦視覺 (CV) 是一個很大的領域，主要關注的是分類問題。這是貓還是狗？X 光片是否顯示了腫瘤？

通過電腦視覺，電腦能夠以更高的速度和效率，以相當於或高於人類的水準準確識別圖像中的人物、地點和事物。它通常採用深度學習模型。CV 自動從單個圖像或圖像序列中提取、分析、分類和理解有用的資訊。圖像資料可能會呈現出多種形式，例如單個圖像、視頻序列、來自多個相機的視圖，或者三維資料。

稍後，您將在本課程中學習有關電腦視覺的更多知識。



下面您將瞭解無監督機器學習。

有時，您所擁有的只有資料 – 沒有監督者。與監督式學習不同，無監督學習沒有人提供標籤，因為所有變數和模式均未知。在這種情況下，機器必須自行發現和創建標籤。這些模型使用提供的資料檢測整個資料集中出現的屬性，然後構建模式。

無監督學習的一個常見子類別是“聚類”。這種演算法根據相似的特徵，將資料分組到不同的聚類，從而更好地瞭解某個聚類的屬性。例如，假設您為一家行銷組織工作，您的組織向為客戶提供雲計算資源的公司提供支援。通過分析客戶的購買習慣，無監督演算法可以識別將特定公司分類（或描述為）小型、中型或大型公司的客戶群。各公司可以利用這些見解，根據自己所處的分類（小型、中型或大型）做出明智的戰略行銷決策，從而更好地滿足客戶對雲計算資源的需求。在這種情況下，集群可以說明您意識到，必須針對不同規模的公司提出不同的行銷策略。

無監督演算法的優勢在於，它們使您可以發現通過其他演算法無從發現的資料模式。例如，存在兩種主要的客戶類型。

自然語言處理 (NLP) 是機器學習的另一個領域，能夠隨著使用的增加不斷發展成熟。例如，Amazon Alexa（或任何其他語音助手）會使用 NLP 嘗試回答您的問題。NLP 與語音和書面文本有關。

NLP 用於許多應用領域，例如：

- 聊天或呼叫中心機器人 – 用於查詢銀行餘額或從餐廳訂餐。
- 翻譯工具 – 即時在各種可翻譯的語言之間進行翻譯，或者提供能即時翻譯功能表的應用程式。
- 語音到文本翻譯 – 將口語轉換為翻譯好的文本。可用于自動翻譯字幕。
- 情緒分析 – 使您能夠分析產品、音樂和電影的評論中的情緒。這些情緒可以用來得出電影的觀眾評分。

您將在後面的模組中學習 NLP。

最近日漸普及的另一種機器學習類型是*強化學習*。與其他機器學習類型不同，強化學習通過挖掘先前反覆運算產生的回饋來不斷改進模型。在強化學習中，代理在環境中進行交互，通過試錯不斷學習。如果知道預期結果的回報，但不知道具體的實現路徑，強化學習非常有用。發現這種路徑需要廣泛試錯。

考慮 AWS DeepRacer 的例子。在 AWS DeepRacer 模擬器中，代理是虛擬汽車，環境是虛擬賽道，行動是汽車油門和轉向輸入，目標是以最快的速度跑完全程，而不偏離賽道。

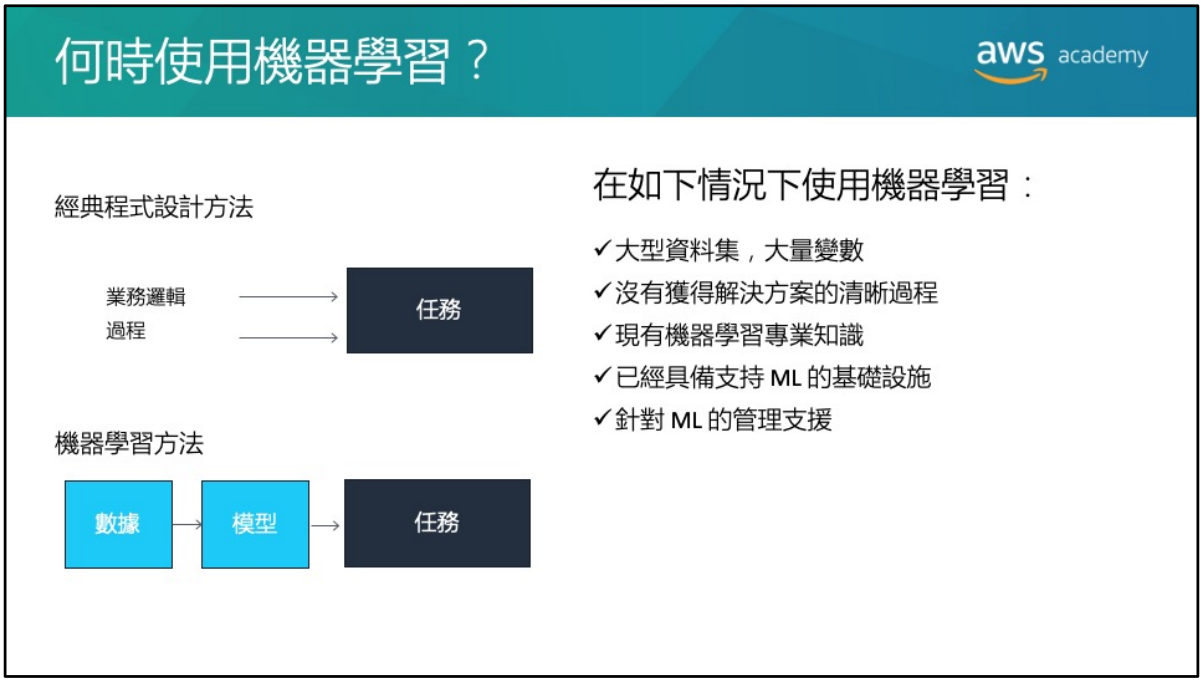
汽車必須學習所需的駕駛行為，以便達到跑完全程的目標。您使用獎勵來激勵模型學習預期的駕駛行為。

在強化學習中，推動學習的元素稱為*代理*。在本例中，代理是 AWS DeepRacer 汽車。*環境*是代理學習所在的場所，在本例中就是帶標記的賽道。行動是指代理在環境中做出了引發回應的*行為*，例如越過了不應該越過的邊界。回應稱為*獎勵*或*懲罰*，具體取決於代理的行動應該是由模型強化的還是阻止的。隨著代理在環境中移動，其行動收到的獎勵應該越來越少，懲罰應該越來越多，直至滿足期望的業務成果。

自動駕駛車輛

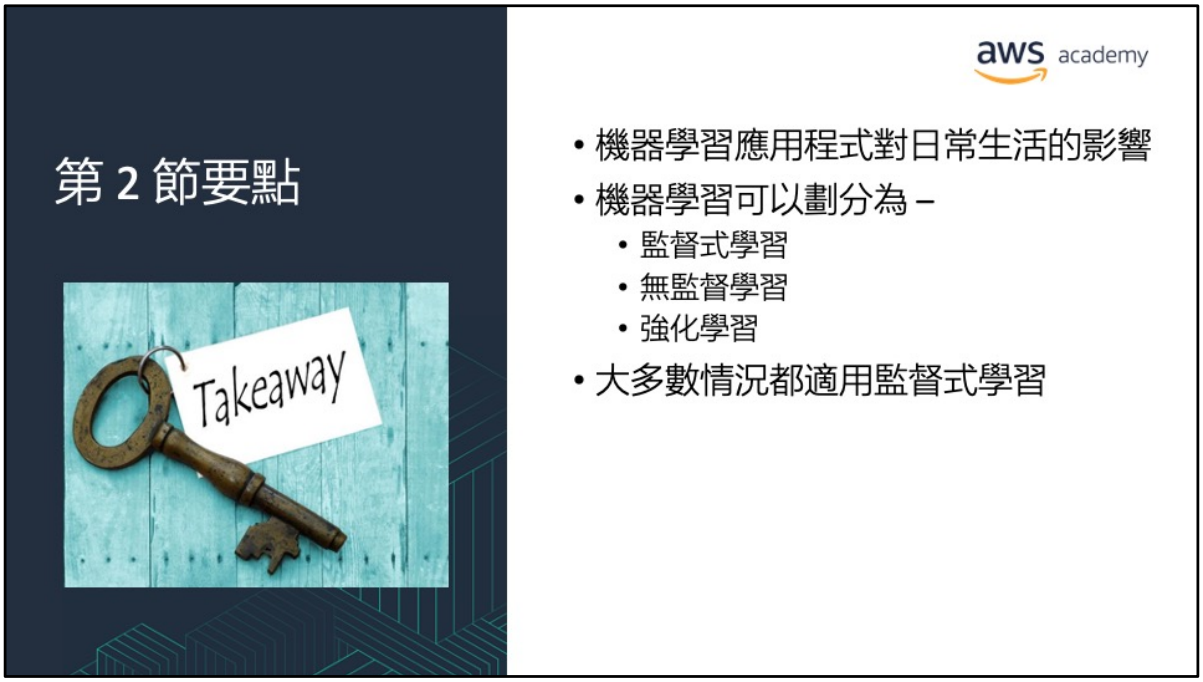
自動駕駛汽車彙集了多種機器學習和深度學習演算法和模型，以解決從 A 點到 B 點的駕駛問題。其中的一項主要任務是不斷檢測環境並預測變化。這項任務涉及到物件檢測，即對檢測到的物件的運動進行定位和預測。這些發現的輸出將作為其他系統的輸入，其他系統會利用這些資訊來決定如何使用車輛的各種控制機制。

一些使用案例涉及到需要對環境進行即時回應的自動駕駛汽車。例如，如果先前看不到的行人從障礙物後面走出來，則必須立即剎車。這類動作不能有任何延誤或錯誤。



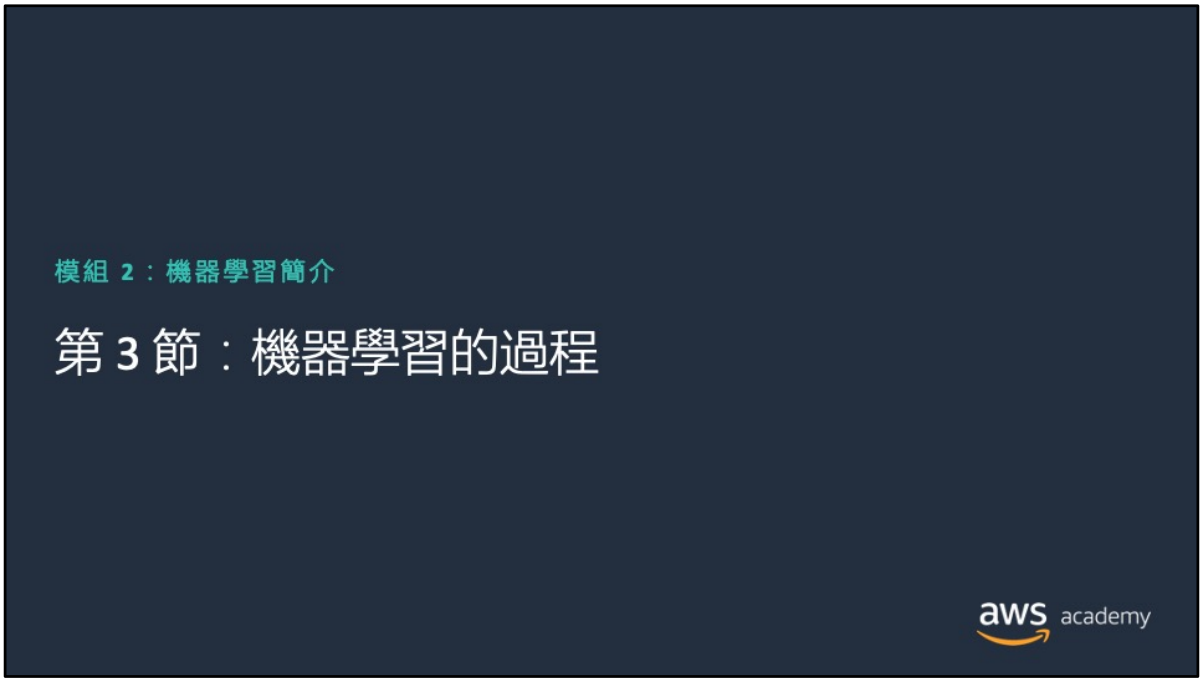
並不是每個問題都應該通過機器學習來解決，有時基礎程式設計也能取得很好的結果。在探索某個情景是否適合採用 ML 解決方案，請注意關注大型資料集和大量變數之類的事物。也有可能您並不確定獲得答案或完成任務所需的業務邏輯或過程。在這種情況下，機器學習通常是最佳選擇。

機器學習系統可能高度複雜。必須具備支持基礎設施、管理支持和專業知識才能幫助專案取得成功。



本模組中這部分內容的要點包括：

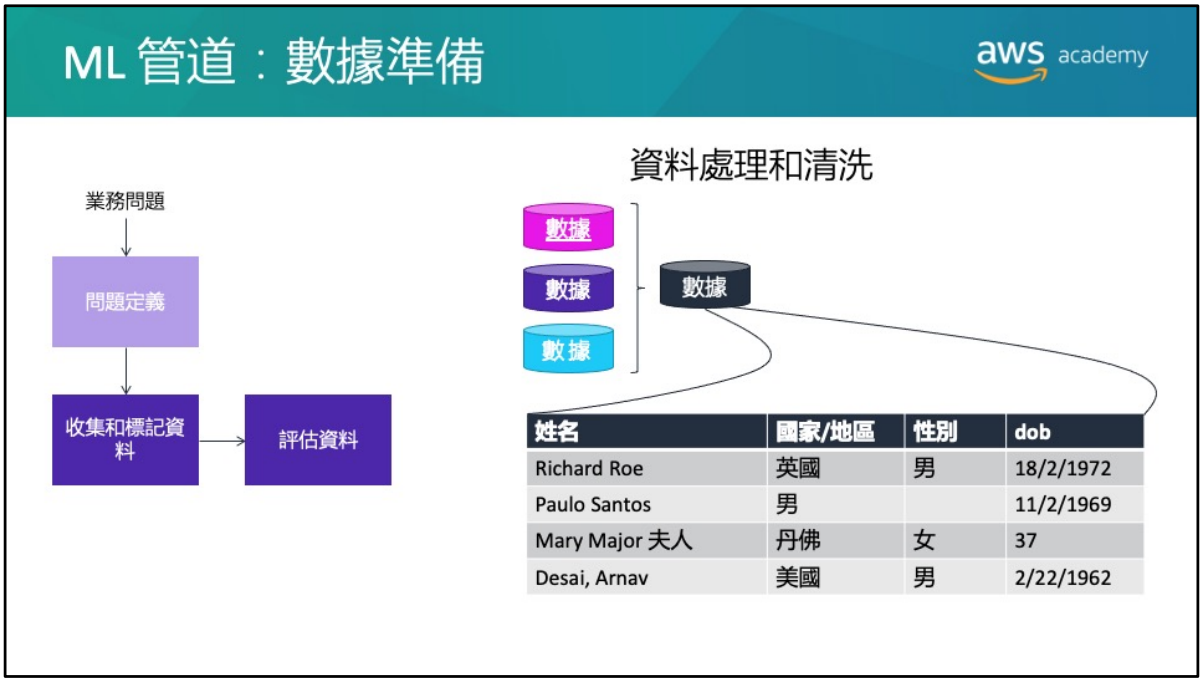
- 機器學習應用程式已經成為人們日常生活的一部分。
- 機器學習客戶劃分為 –
    - 受監管學習：您針對您知道的答案進行資料訓練。
    - 無監督學習：您有資料，但您正在尋找對資料深入的見解。
    - 強化學習：模型通過基於經驗和回饋的方式學習。
- 大多數業務問題是監督式學習問題。



介紹“第 3 節：機器學習的過程”。

在本節中，您將獲得有關機器學習術語和典型工作流程的簡要概述。在後續課程中您會瞭解有關該主題的更多詳細資訊。

首先，您應該始終從您或您的團隊認定可以得益於 ML 的業務問題入手。然後，您需要定義問題。在此階段，您需要闡明您的業務問題並將其轉化為 ML 問題。

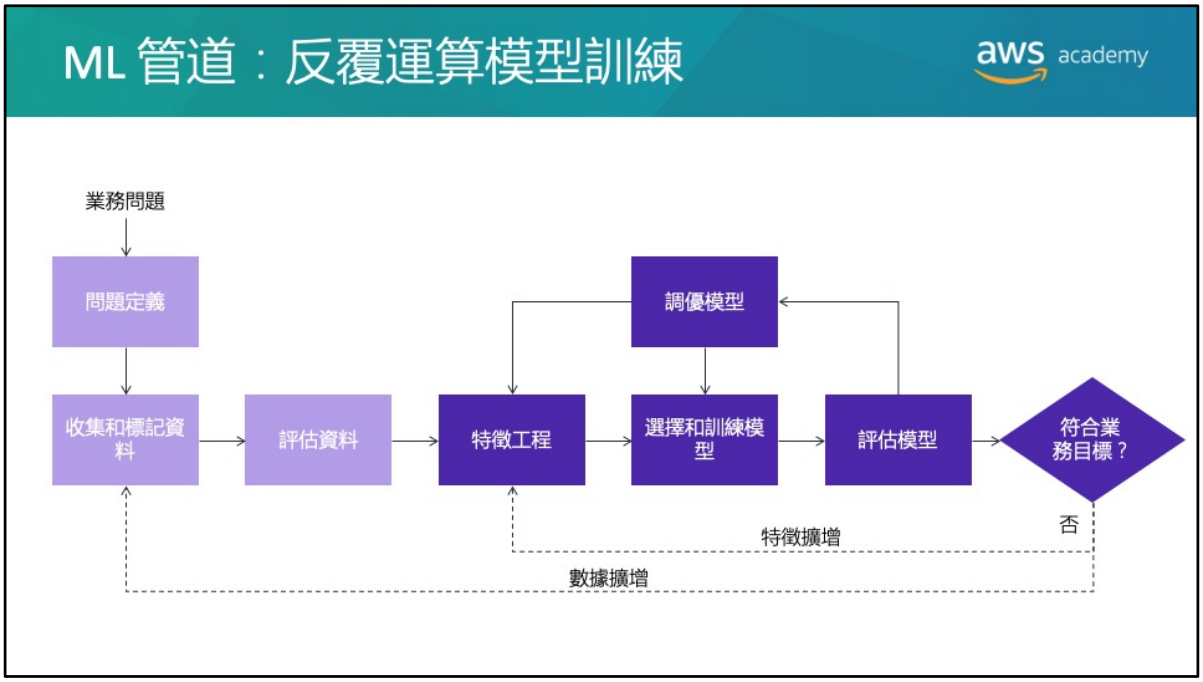


定義問題後，您要進入資料準備和預處理階段。在此階段中，您將從一個或多個資料來源中提取資料。這些資料來源的資料或類型可能有所不同，必須對這些資料或類型進行協調才能形成資料的統一視圖。您必須為資料添加視覺化效果，並使用統計學來確定資料是否一致並且可用於機器學習。在後續課程中，您將查看一些資料來源。

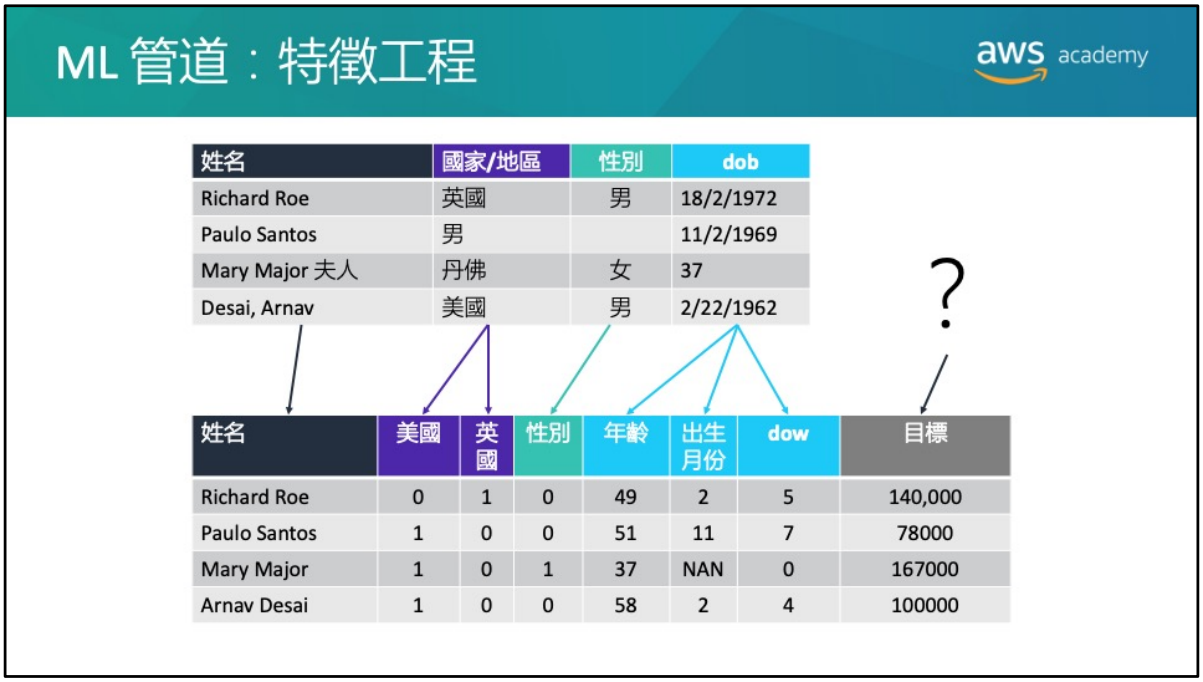
示例資料包含四列，其中包含著從三個資料來源收集的資料。各個資料來源表示資料的方式略有不同，結果顯示在表中。在 ML 問題中，列表示特徵，而行代表實例。您可能發現，部分實例中的資料存在問題。

在某些情況下，您需要主題專家或職能專家的說明來瞭解資料的真實性。例如，*11/2/1969* 可能表示 *11 月 2 日*，也可能表示 *2 月 11 日*。擁有或管理相應資料池的人將能夠闡明該資料的含義。*Male* 或許可以歸因於在導入時儲存格移位而造成的問題，但也可能代表地理位置：馬爾地夫共和國首都馬累。有時，這種錯誤識別並不那麼簡單，需要 SME 的說明。您將在稍後的課程中瞭解專家的作用。

通過獲得一致且正確的資料，您可以對 ML 專案的成功產生重大影響。



準備好資料並確保其正確性和一致性之後，就該對模型進行訓練了。在此階段，過程變為反覆運算、流動式。在找到可以滿足業務目標的模型之前，您可能需要經歷多次特徵工程、訓練、評估和調優。

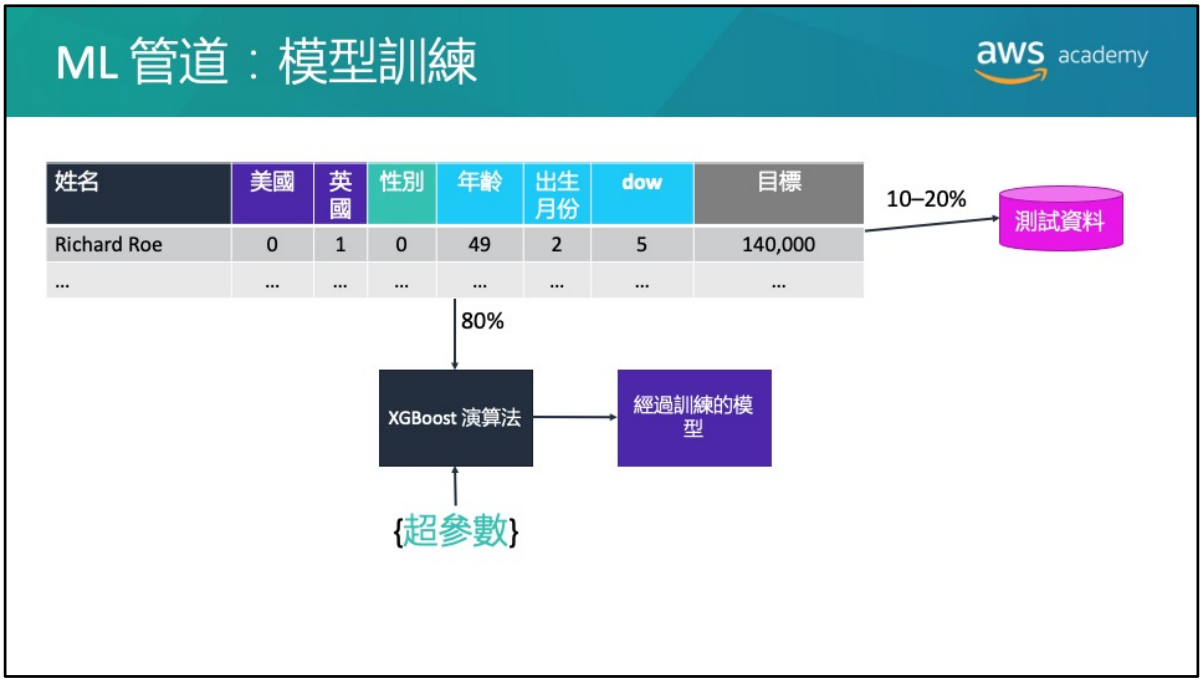


*特徵工程*是指選擇或創建特徵的過程，您將使用這些特徵訓練模型。*特征*是您的資料集中已有的資料列。模型的目標是嘗試準確估計新資料的目標值。ML 演算法使用特徵來預測*目標*。在此示例中，目標資料是指一周內執行步驟的平均數量。

選擇正確的特徵可能包括添加、刪除或計算新特徵。在特徵工程期間，您可能會遇到以下更改：

- **“Name”（姓名）列的資料清理** – 您可能想要讓格式保持一致，這在模型稍後的應用中會很有用。您也可能會僅出於美觀原因更改格式。根據您要使用此資料解決的問題，甚至可能不需要 *Name*（姓名）。
- **將國家/地區轉換為數位或二進位列集** – 如果此資料庫是傳統資料庫，則可能需要將 *Country*（國家/地區）移至查閱資料表，然後對其進行引用。大多數 ML 演算法都希望實例的資料以單行呈現。此外，ML 演算法需要處理數值資料。您可以考慮將“國家/地區”文本轉換成國家/地區 ISO 代碼。但是，該模型可能會將數值解釋為某種含義，因此 *UK (44)* 比 *USA (01)* 更有意義。在這種情況下，就可以將資料拆分成多個列。這種做法稱為*分類編碼*，您將在本課程的後面部分中學習相關知識。
- **將性別資料轉換為二進位數字** – 如果將文本值轉換為數值，即 *0* 或 *1* 代表*男*或*女*，則模型可以更輕鬆地使用它。
- **將出生日期 (DOB) 拆分成多個組成部分** – 提取年齡、出生月份 (*bm*) 和星期幾 (*dow*) 可能是恰當的做法，具體取決於您嘗試解決的問題。年齡是否會影響目標變數？生日那天是星期幾呢？

如果這些資訊看起來十分複雜，您不必擔心。您將在本課程的後面部分瞭解有關特徵工程的更多詳細資訊。

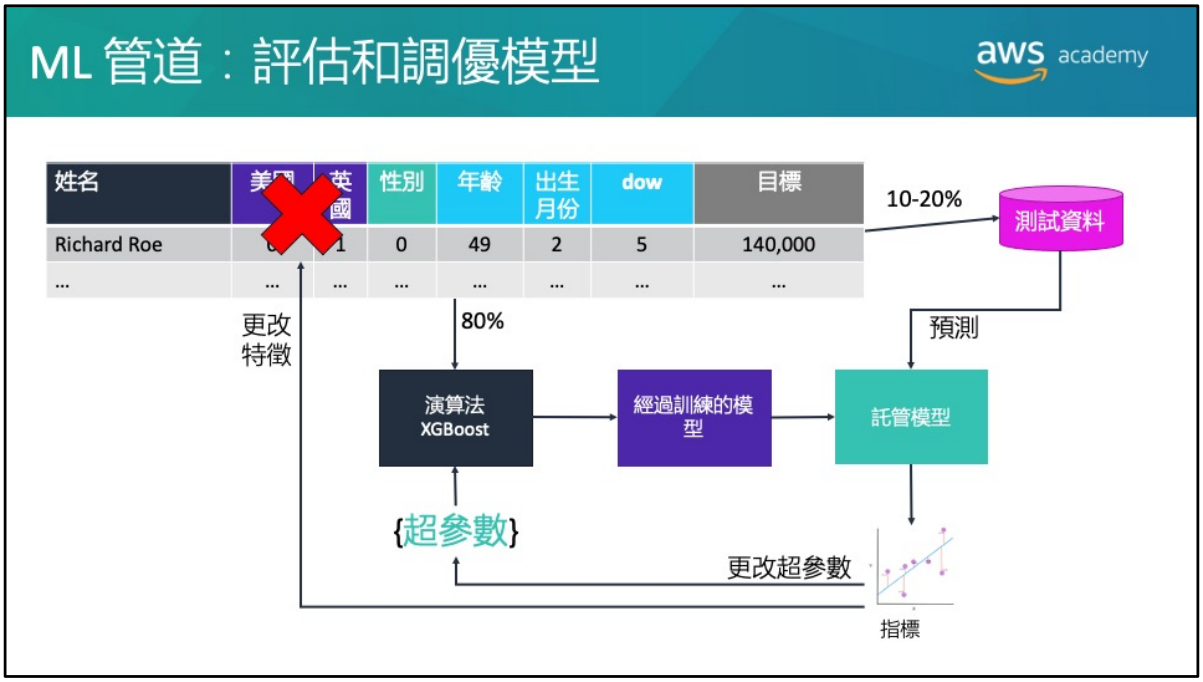


在清理好資料並確定其特徵後，即可開始訓練模型。您無需使用所有資料來訓練模型。相反，您應該保留一些資料，以便可以有一些資料用來執行測試。通常情況下，您可以使用 80% 的資料用於訓練，將其餘部分留作測試使用。

您可以使用訓練資料來對模型進行訓練。在圖中，所用模型是 XGBoost 演算法。

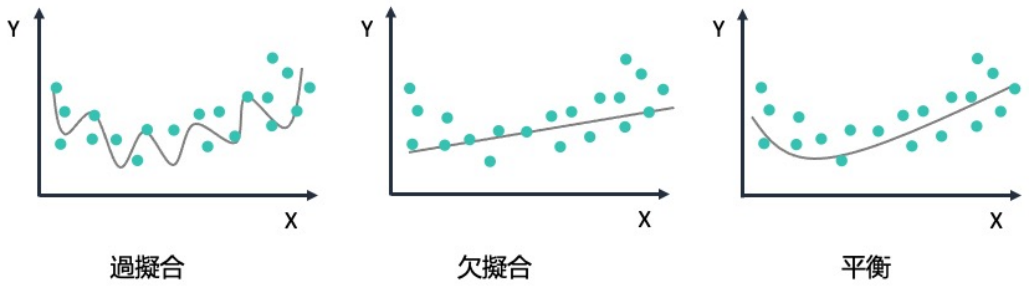
該模型本身具有一些您可以設置的參數，它們會改變演算法的工作方式。這些參數稱為*超參數*。

訓練作業的輸出是經過訓練的模型。



借助經過訓練的模型，您可以使用一些測試資料瞭解模型的運行效果。您可以選擇一個該模型從未見過的實例，並使用它來執行預測。因為您已經知道測試資料中的目標，所以可以對兩個值進行比較。通過這些比較，您可以計算指標，從而瞭解有關模型執行效果的資料。然後，您可以修改模型的資料、特徵或超參數，直到找到產生最佳結果的模型為止。

過擬合和欠擬合



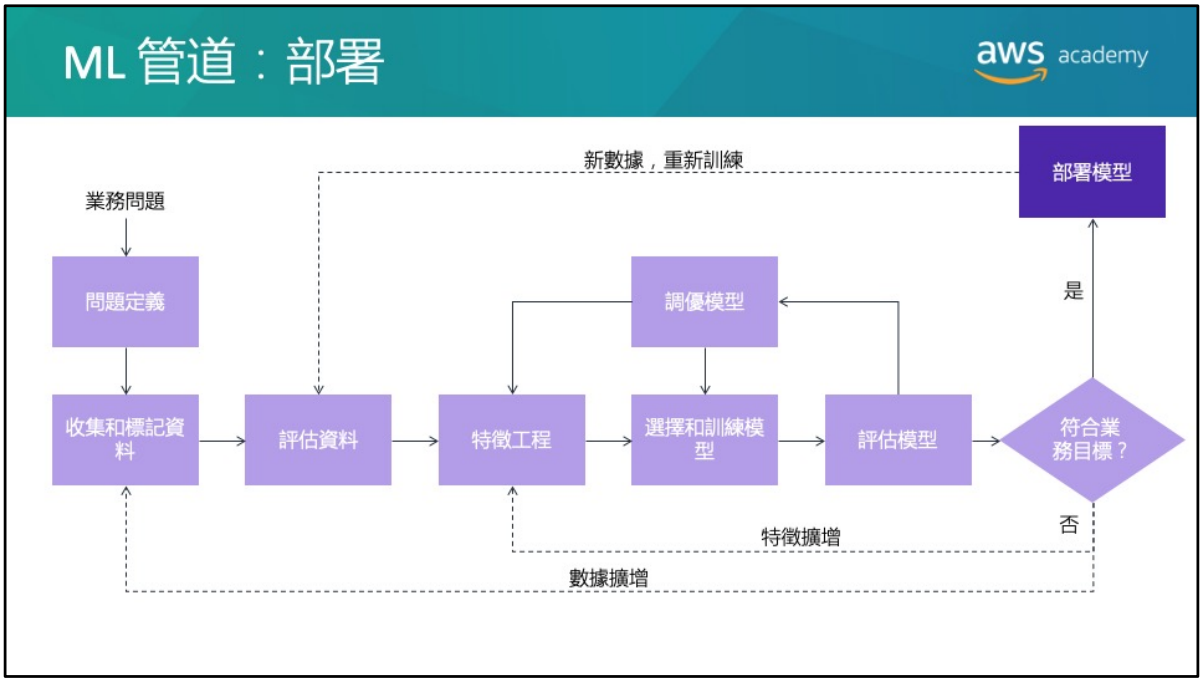
在您訓練模型時，應該認識到過擬合或欠擬合模型的危險。

如果您的模型在訓練資料上表現良好，但在評估資料上表現不佳，則表示您的模型與訓練資料*過擬合*。該模型會存儲它所見到的資料，但無法將其一般化到未曾見過的示例中，這會導致過擬合。

模型在訓練資料上性能欠佳時，表示您的模型與訓練資料*欠擬合*。模型無法捕獲輸入示例（通常稱為 *X*）與目標值（通常稱為 *Y*）之間的關係。這種能力不足會導致欠擬合。

瞭解*模型擬合*對於瞭解模型準確率欠佳的根本原因至關重要。此概念會指導您採取糾正措施，以確定預測模型與訓練資料是欠擬合還是過擬合。這些步驟涉及瞭解訓練資料和評估資料上的預測誤差。

在本課程的稍後部分，您將學習可以避免該問題的步驟。



在重新訓練模型並對結果感到滿意之後，即可部署模型以提供最佳預測結果。

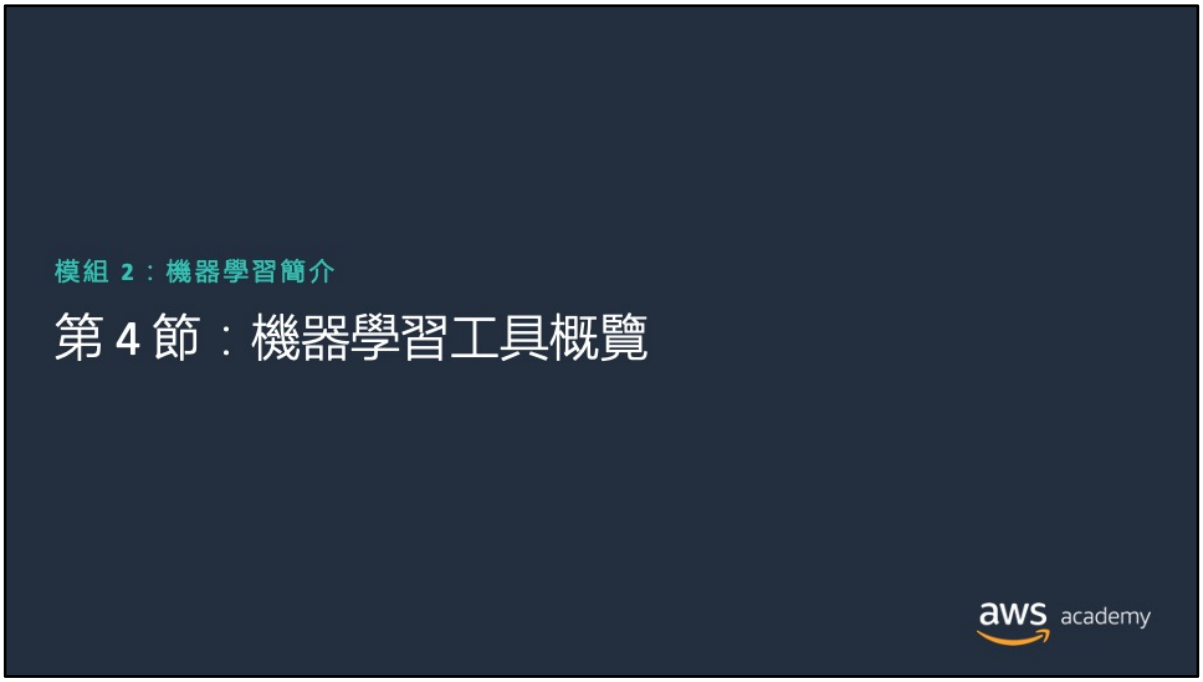
在本課程稍後部分，我們將指導您完成這些不同的階段，並實際動手體驗各個階段。在您使用要探索的託管服務時，瞭解該過程也非常有用。但是，Amazon ML 服務會為您完成工作中最艱難的部分。

第 3 節要點

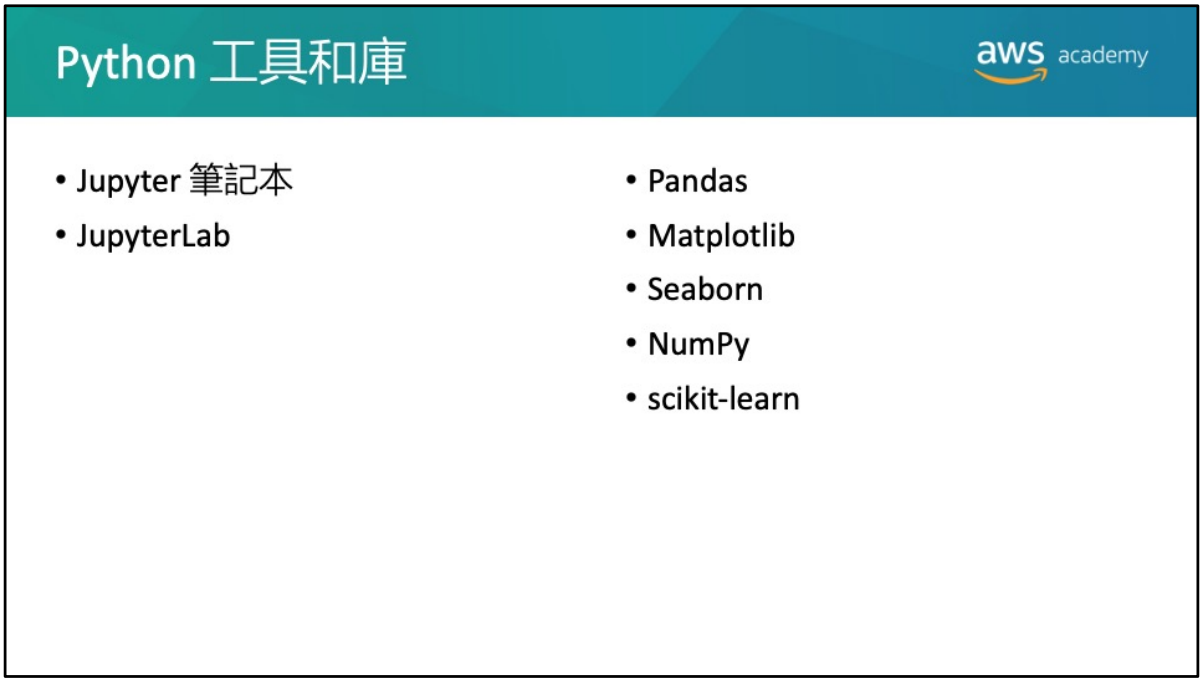
- 機器學習管道會指導您完成機器學習模型評估和訓練
- 反覆運算式流程 -
  - 資料處理
  - 訓練
  - 評估

本模組中這部分內容的要點包括：

- 機器學習管道過程可以指導您完成訓練和評估模型的過程。
- 反覆運算流程可以分為三個主要步驟 –
  - 數據處理
  - 模型訓練
  - 模型評估



介紹“第 4 節：機器學習工具概覽”。



現在，您將瞭解一些用於機器學習的工具。注意：此清單中未列出當前可用的全部 ML 工具，在本課程中您只會用到其中的少數幾種。

*Jupyter 筆記本*是一種開源 Web 應用程式，可用於創建和共用包含即時代碼、方程式、視覺化效果和敘述文本的文檔。用途包括：資料清除和轉換、數位類比、統計建模、資料視覺化和機器學習等。

JupyterLab 是用於 Jupyter 筆記本、代碼和資料的一種基於 Web 的互動式開發環境。JupyterLab 非常靈活。您可以配置和安排使用者介面，以支援資料科學、科學計算和機器學習中的各種工作流程。JupyterLab 還具有可擴展和模組化的特點。您可以編寫外掛程式，從而添加新元件並與現有元件集成。

在本課程的稍後部分，您將使用 Amazon SageMaker，它託管 Jupyter 筆記本和 JupyterLab。

Pandas 是一個開源 Python 庫。它用於資料處理和分析。它以類似於試算表的表格形式表示資料。該表格稱為 *Pandas DataFrame*。

Matplotlib 是一個庫，用於在 Python 中創建靜態、動畫式和互動式的科學視覺化內容。在隨後的課程中，您將用它來生成資料圖。

Seaborn 是另一個 Python 資料視覺化庫。它基於 matplotlib 構建，並提供了用於繪製內容豐富的統計圖形的高級介面。

NumPy 是 Python 中的一個基礎科學計算包。它包含用於 N 維陣列物件的函數和大量有用的數學函數，例如線性代數、傅立葉轉換和亂數功能。

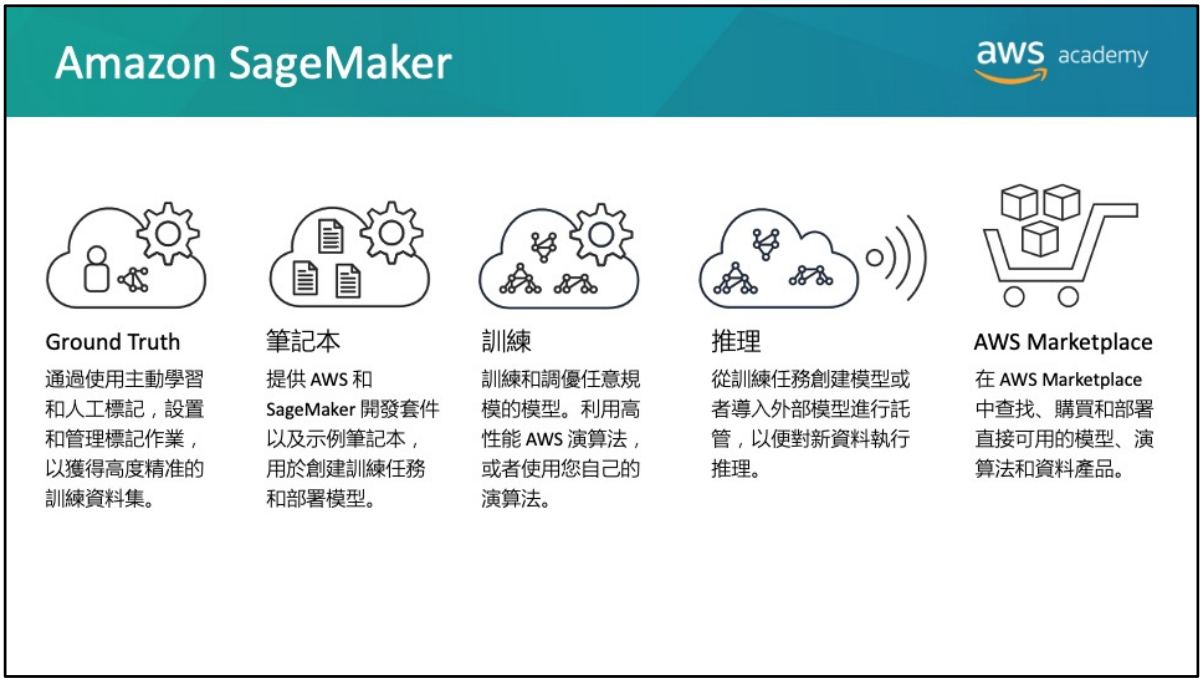
scikit-learn 是一種開源機器學習庫，它支持監督式學習和無監督學習。它還提供了用於模型擬合、資料預處理、模型選擇和評估的工具以及許多其他實用工具。scikit-learn 基於 NumPy、SciPy 和 matplotlib 構建，是探索機器學習的理想工具包。儘管在後面的模組中，您僅會使用這一工具包的少數幾個函數，但是在完成本課程後，您可以更詳細地研究此工具包。

除了單個的庫和工具包之外，您可以使用包含適用於生產環境的框架的工具。

您已經簡要瞭解了 scikit-learn，這是一種非常適合學習的庫。您還可以使用其他庫，例如用於機器學習的 TensorFlow 和用於深度學習的 Keras。

AWS 支持所有這些框架，並且可以通過 Amazon SageMaker 使用所有這些框架。

此外，Amazon 還在雲端和邊緣提供了針對機器學習調優的計算實例。計算實例針對學習和推理進行過調優，並預先打包了包含許多流行框架的 Amazon Machine Images (AMI)。



最後，您可以使用 Amazon SageMaker，這是一項功能繁多的 AWS 服務。

Amazon SageMaker 可以部署運行 Jupyter 筆記本和 JupyterLab 的機器學習實例。Amazon SageMaker 管理這些計算資源的部署，因此您必須連接到 Jupyter 環境。Amazon SageMaker 還提供了用於標記資料、訓練模型和託管訓練模型的工具。AWS Marketplace 提供了面向機器學習開發人員的精選現成模型包和演算法。

您的講師現在將演示 Amazon SageMaker 或為您提供演示錄影的存取權限。

AWS 提供了一組託管 ML 服務，即便您沒有 ML 經驗，也可以將它們集成到自己的應用程式中。

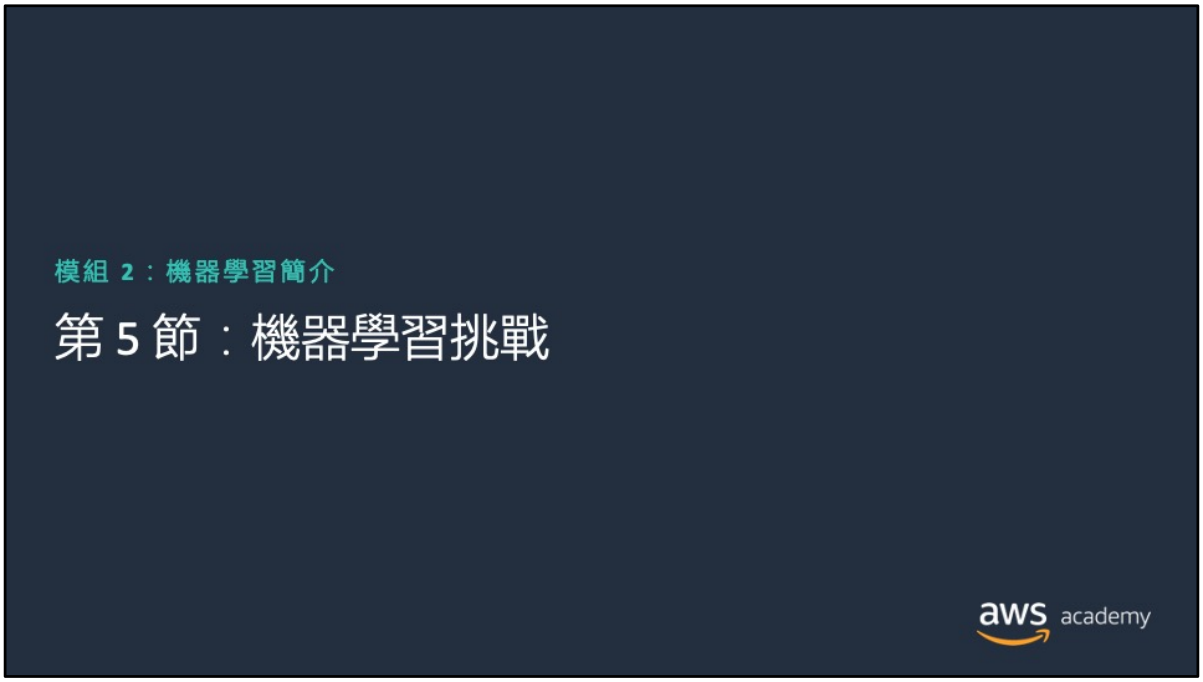
- 計算視覺 – Amazon Rekognition 為圖像和視頻提供物件和人臉識別功能。Amazon Textract 可以從圖像中提取文本。
- 語音 – Amazon Polly 可以朗讀文本，而 Amazon Transcribe 可以將語音轉換為文本。
- 語言 – Amazon Comprehend 使用 NLP 在文本中發現見解和關係。Amazon Translate 可以將文本翻譯為不同語言。
- Chatbot – Amazon Lex 服務可說明您構建使用語音或文本的互動式對話應用程式。
- 預測 – Amazon Forecast 使用機器學習將時間序列資料與其他變數相結合，以構建預測。
- 推薦 – Amazon Personalize 是另一項機器學習服務，可以說明您為客戶創建個性化的推薦建議。

這些託管服務在相關問題域的許多方面經過訓練，您可以提供特定的資料來啟動該過程。

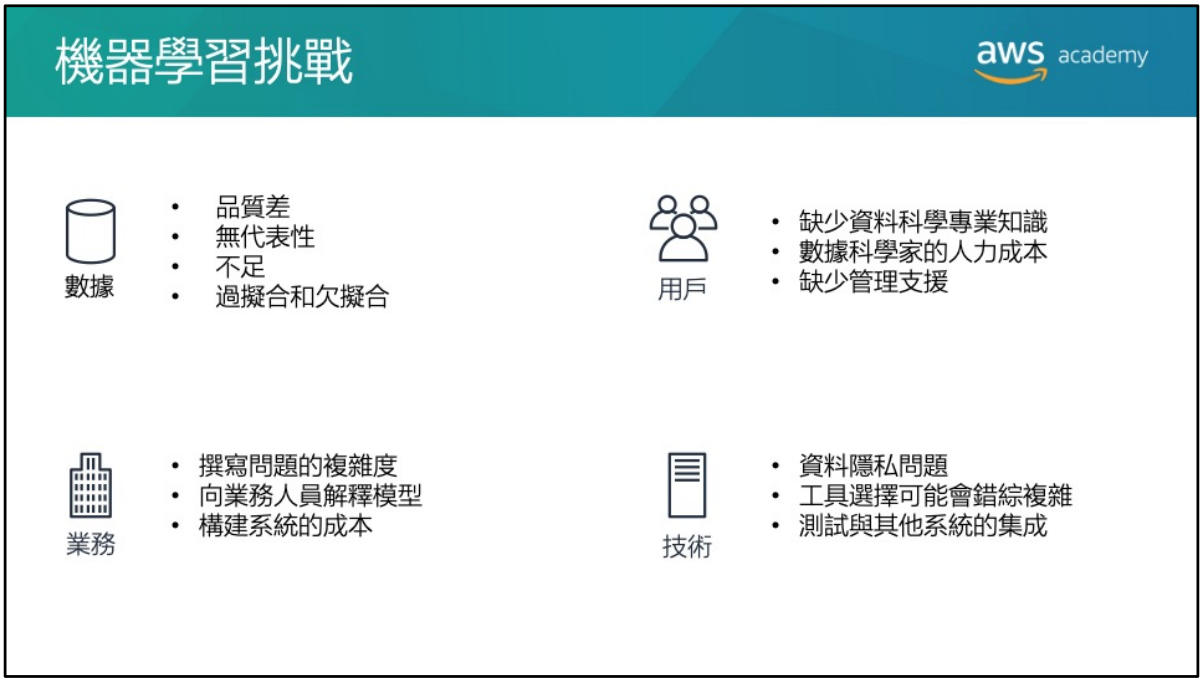
在學習了如何自行執行許多操作之後，您將在本課程的後半部分中學習其中許多託管服務。

本模組中這部分內容的要點包括：

- Python 是執行機器學習任務時最受歡迎的語言。
- Jupyter 筆記本為您提供了一個由 Web 流覽器託管的機器學習開發環境。
- 目前有大量的開源工具，本課程中僅介紹了少數重要工具。
- 根據您的需求，您可以開始使用低級別框架，也可以使用您自己的解決方案。您可以使用 Amazon SageMaker 之類的工具來幫助完成更大規模的任務，或者針對您的特定問題域調整一種託管 ML 服務。



介紹“第 5 節：機器學習挑戰”。



在機器學習中，您會遇到許多挑戰，如以下示例所示。

數據：
- 世界上存在大量品質差、不一致的資料。您的工作的重點就是獲取良好的資料。
- 數據是否很好地代表了問題？如果您嘗試查明信用卡欺詐，那麼是否有可用於訓練的相應示例？
- 您擁有的資料量是否充足？大多數情況下，您擁有的資料越多越好。
- 您的模型是否過擬合或欠擬合？

用戶：
- 您是否有資料科學方面的經驗？
- 數據科學家團隊的人員配置是否經濟高效？
- 管理層是否支援使用 ML？

業務：
- 問題是否過於複雜，難以制定 ML 問題？
- 能否向業務部門解釋所得到的模型？如果不能，那麼這種模型可能就無法得到採用。
- 構建、更新和運行 ML 解決方案的成本是多少？

技術：
- 業務部門能否訪問您需要的資料？能否對資料實施保護，以滿足法規要求？
- 您計畫使用哪些工具和框架？
- 此解決方案如何與其他系統集成？

如今，通過利用現有模型即可在所掌握的 ML 知識不多的情況下解決許多 ML 問題。

您已瞭解了用於 ML 的 AWS Managed Services。現在，您需要一些 API 調用方面的開發人員技能，並且可以向應用程式添加複雜的機器學習功能。

您可以使用其他預先構建模型，也可以基於廣受歡迎的視覺模型 You Only Look Once (YOLO) 做出調整。

或許您並不想創建自己的演算法，而是想購買獨立軟體廠商開發的模型和服務。除了先前介紹的情景之外，您還可以通過 AWS Marketplace 查找協力廠商解決方案。

第 5 節要點

- 機器學習挑戰
  - 數據
  - 人員
  - 業務
  - 技術
- 托管服務可簡化機器學習

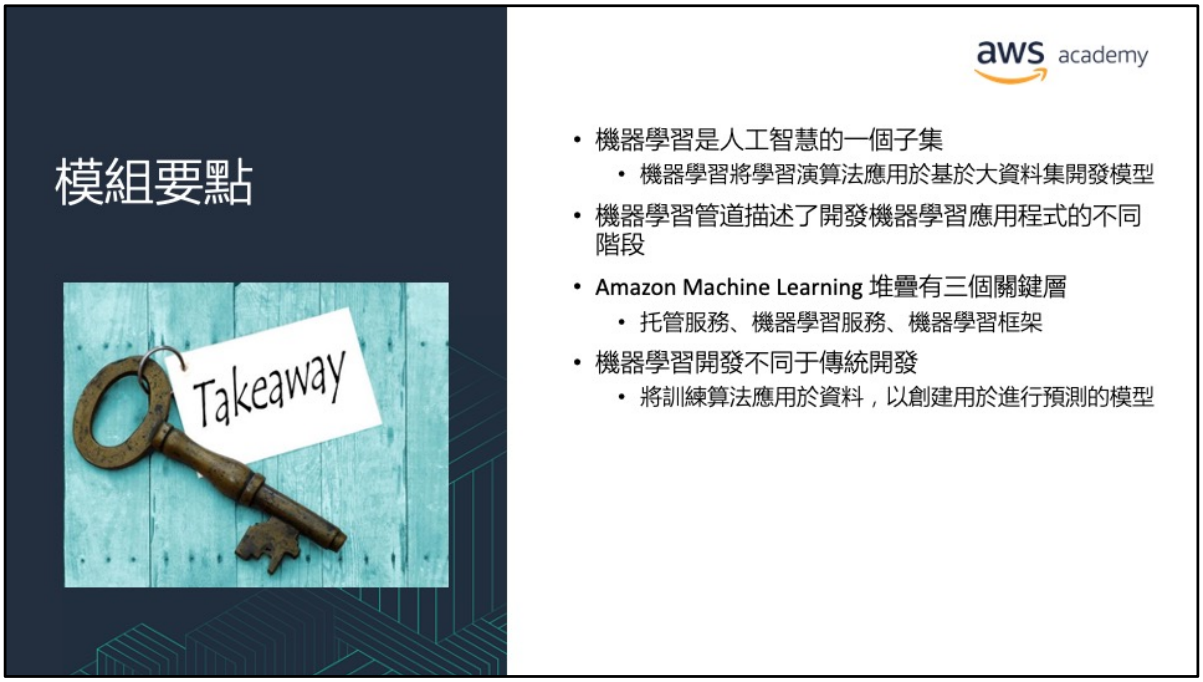
本模組中這部分內容的要點包括：

- 您會遇到許多機器學習難題。您能直接產生影響的最大問題與資料有關，但是您還要處理人員、業務和技術挑戰。
- 不妨考慮使用託管服務來說明您解決機器學習問題。

模組 2：機器學習簡介

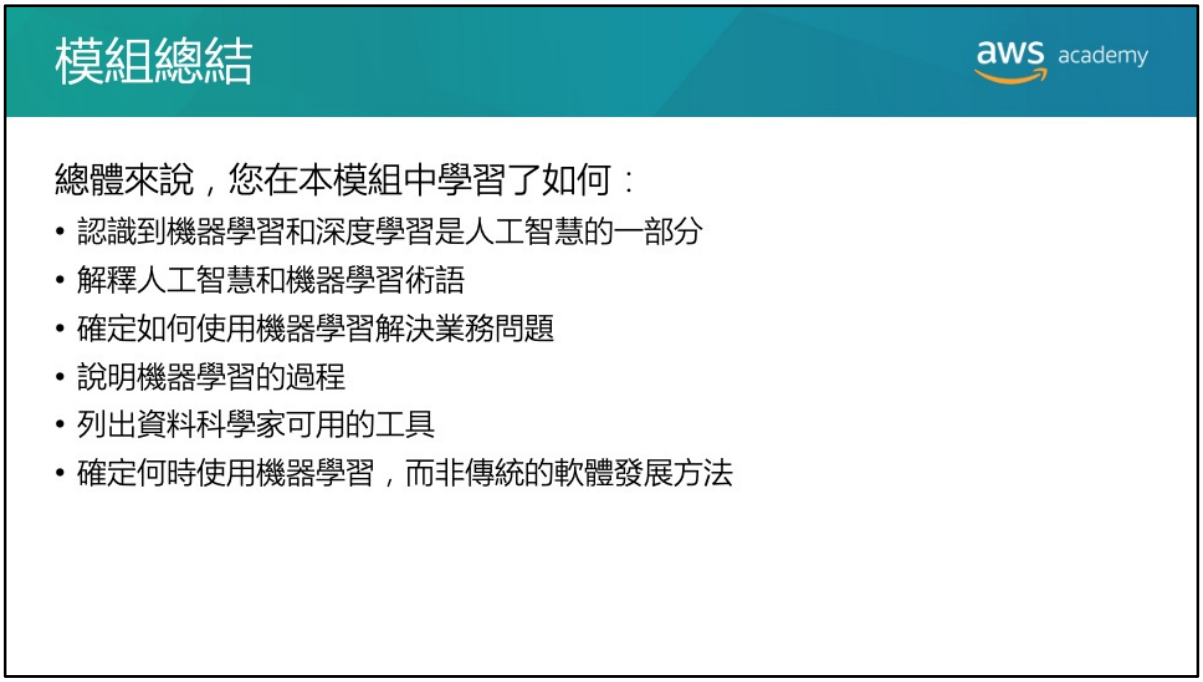
模組總結

現在，我們來回顧和總結一下本模組，然後進行知識測驗。



本模組中這部分內容的要點包括：

- 機器學習是人工智慧的一個子集。
    - 機器學習將學習演算法應用於基於大資料集開發模型。
- 機器學習管道描述了開發機器學習應用程式的不同階段。
- 機器學習堆疊具有三個關鍵層。
    - API 服務、機器學習服務和機器學習框架
- 機器學習開發不同于傳統開發。
    - 它會將訓練演算法應用於資料，以創建用於進行預測的模型。



總體來說，您在本模組中學習了如何：

- 認識到機器學習和深度學習是人工智慧的一部分
- 解釋人工智慧和機器學習術語
- 確定如何使用機器學習解決業務問題
- 說明機器學習的過程
- 列出資料科學家可用的工具
- 確定何時使用機器學習，而非傳統的軟體發展方法

完成知識測驗

現在可以完成本模組的知識測驗。

如果您想瞭解有關本模組所涵蓋主題的更多資訊，下面這些其他資源可能會有所幫助：

- 什麼是機器學習？
- AWS 上的機器學習


謝謝
© 2020 Amazon Web Services, Inc. 或其附屬公司。保留所有權利。未經 Amazon Web Services, Inc. 事先書面許可，不得複製或轉載本文的部分或全部內容。禁止因商業目的複製、出借或出售本文。如有對本課程的糾正或回饋意見，請發送電子郵件至：aws-course-feedback@amazon.com。如有其他任何問題，請與我們聯繫：https://aws.amazon.com/contact-us/aws-training/。所有商標均為各自所有者的財產。
aws academy

感謝您的參與！