

01

何謂人工智慧？

在 1956 年的達特茅斯會議（Dartmouth Conference）上，人工智慧一詞首度被提出。會議中，人們討論了電腦進行智慧資訊處理的議題。經過半世紀以上的現在，我們該如何定義人工智慧呢？

○ 定義模糊的人工智慧

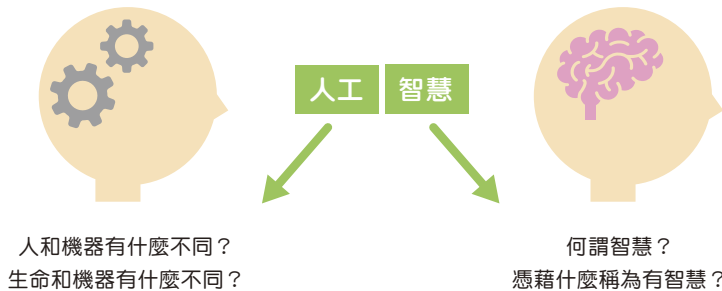
人工智慧（Artificial Intelligence）沒有辦法簡單定義。

首先，「人工」一詞會衍生疑惑：人類和機器的區分基準為何？而「智慧」一詞會產生疑問：憑藉什麼稱為有智慧？我們必須回答這兩個疑問，才有辦法定義人工智慧。

雖然在前線活躍的研究員提出了各種答案，但目前仍舊沒有明確的定義。

因此，我們可先大致定義為「能夠如同人類進行智慧處理的技術、機器」，再根據不同的用途來學習相關的術語。

■ 人工智慧的定義



能夠如同人類進行智慧處理的技術、機器？

○ 人工智慧的分類方法

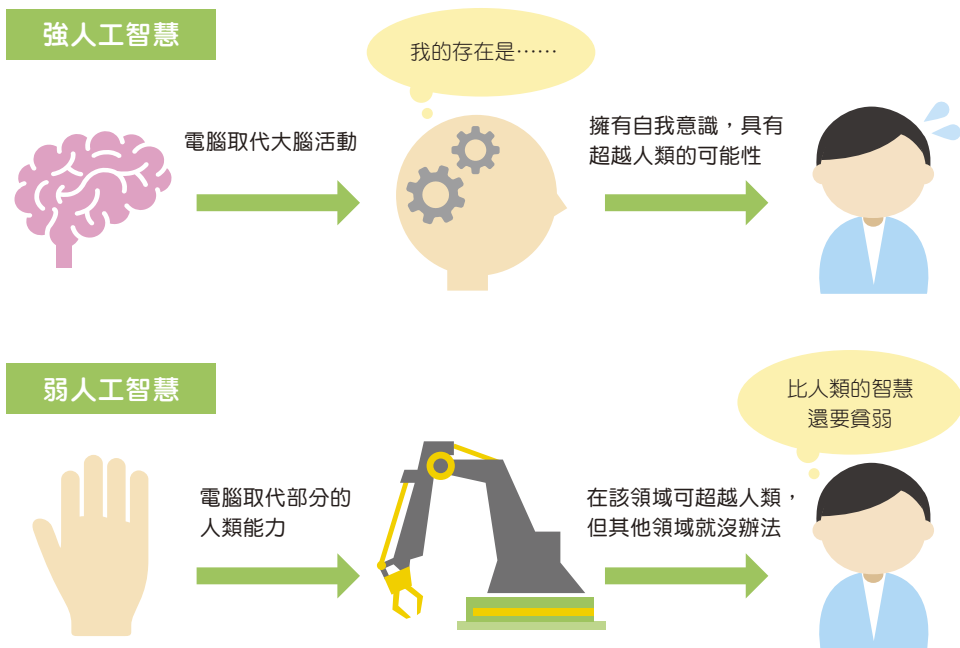
雖然人工智慧難以定義，但存在幾種分類方式。

其中一種分類是，哲學家約翰·希爾勒（John Searle）提出的「**強人工智慧**」和「**弱人工智慧**」，著眼於人工智慧的認知狀態來區分。

強人工智慧是指，以模仿智能本身，獲得如同人類認知狀態的機器，比如，哆啦 A 夢、原子小金剛等漫畫人物。強人工智慧具有壓倒性運算能力，被認為將帶來機器超越人類的技術奇點（Technological Singularity）。

而弱人工智慧是指，以模仿人類（有智慧）的行動，獲得部分人類能力的機器，比如，下將棋、黑白棋的電腦，與後面會詳細介紹的圖像辨識等。雖然這類人工智慧的行為看起來具有智慧，但本身並沒有自我存在的意識。

■ 強人工智慧與弱人工智慧



02

何謂機器學習 (ML)?

機器學習是人工智慧的分類之一，為幫助電腦有效率、有效果學習的理論體系。它只需進行適當的處理，就能根據輸入資料預測數值、最佳化，因此目前已經活用於各式各樣的領域。

○ 人工智慧的關鍵 —— 機器學習

為了讓電腦具有更高水準的辨識能力，必須決定根據什麼基準來運作，而這個基準稱為**參數**。比如，某人工智慧根據身高判斷人類圖像是大人還是小孩，此時的身高就是參數。機器學習能夠根據輸入的資料，自動決定（學習）最為正確的行為，因而被視為人工智慧發展的關鍵。

在機器學習以前，主流是記住全部資料的記憶學習，但這種作法無法解答未知的資料。然而，近年隨著資訊技術的進展，變得能夠低成本獲得、累積名為**大數據**的大量資料。使用大數據反覆訓練後，即便輸入未知的資料也能得到解答。雖說如此，現今的機器學習未必都是利用大數據。

■ 與記憶學習迥異的機器學習

	分數	合格
A	100	○
B	90	○
C	80	○
D	70	×
E	85	?

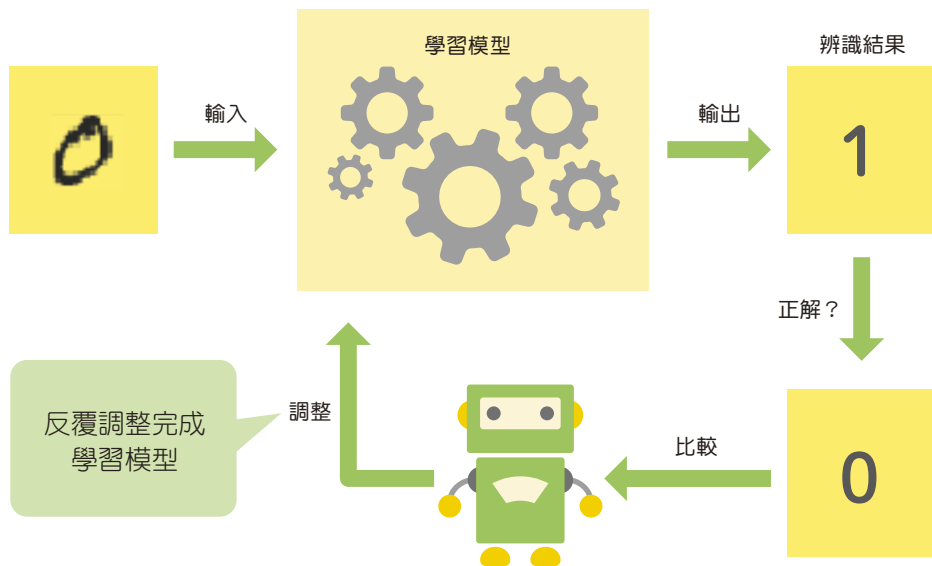
機器學習的程序

機器學習是電腦接收輸入資料後，使用學習模型輸出運算結果。**學習模型**相當於人工智慧的大腦，輸入某資料後，輸出更為適當的決策資料。

第一道程序是，比較預期的輸出資料（標籤、監督訊號）與學習模型的運算結果，來修正學習模型。經過反覆修正後，保存最終的學習模型，結束學習的處理。另外，學習模型也可單獨稱為「模型」。

根據上述內容來分類手寫數字，準備好大量手寫數字（0～9）的圖像資料，圖像與其表示的數字（正解）兩兩成對。將這些資料輸入學習模型，起初會輸出荒唐的數值，輸入0的手寫數字圖像有可能輸出成1。比較荒唐的數值與正解數字修正學習模型，經過不斷反覆修正，輸出值會逐漸趨近正確的數字。完成學習模型後，讓它讀取手寫數字，利用輸出值進行數字的圖像辨識。

機器學習的圖像辨識過程



05

監督式學習的機制

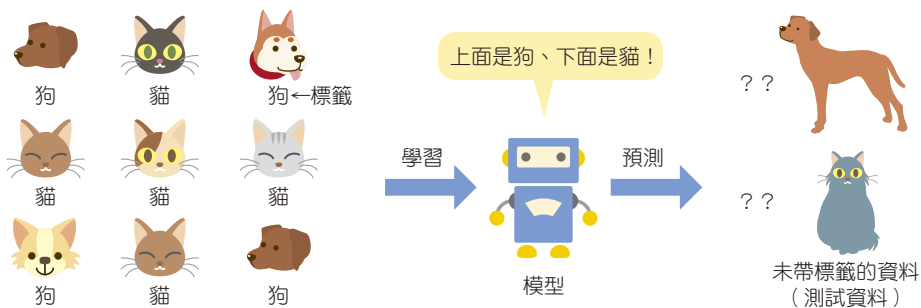
機器學習之一的監督式學習，因實際執行的處理與名稱意象接近，而較為容易理解。如同其名，這是透過資料的標籤發揮監督的功用，由人類教導機器範例的手法。

○ 何謂監督式學習？

監督式學習 (supervised learning) 是指，使用已有正確解答的資料讓模型學習的方法。這邊所說的模型相當於人工智慧的大腦部分，正解的答案稱為**標籤**；含有解答的資料稱為**附帶標籤的資料 (或者訓練資料)**。雖然監督式學習是使用附帶標籤的資料進行模型的學習，但最終目標是能夠正確解答未帶標籤的資料 (測試資料)。Section02 的機械學習例子屬於監督式學習。

比如，使用監督式學習解決貓狗圖像的分類問題。人類事前對圖像分別標註貓或狗的標籤，模型觀看圖像與標籤的對應關係，學習哪張圖像是貓、哪張圖像是狗。如果最後即便沒有貓狗標籤，也能夠判斷為何種圖像，則代表模型有順利學習。

■ 特徵量的作法



分類與迴歸

監督式學習可分為分類與迴歸兩種類型，分類有時又被稱為「識別」。在 Section02 提到，分類是「以盡可能區分整個資料來畫線」；迴歸是「以盡可能重疊整個資料來畫線」。本小節會以「被預測的數值（答案）為何？」的角度，來解說分類與迴歸的不同。

首先，分類的答案會是「貓／狗」、「小學生／國中生／高中生／大學生」等不同的類別。這邊所說的類別，滿足 ① 不是連續數值（**離散值**）、② 大小、順序沒有意義。然而，即便乍看之下答案是連續的數值，若該數值可視為不同的類別（離散值），則屬於分類識別的問題。比如，推測一位數的手寫數字為何的分類問題，答案會是 0/1/2/3/4/5/6/7/8/9 的其中之一，而不是 0.5、2.1 等沒有意義的答案。再則，辨識圖像時，僅會關注辨識的數字結果是否正確，而不會在意數字的大小關係。因此，答案可視為不同的類別。

另一方面，迴歸的答案會是連續的數值（**連續值**）。討論股價的預測問題時，即便答案是 12345.6 日圓這種不上不下的數值，也具有意義。因此，股價的預測問題會屬於迴歸。

■ 分類與迴歸

分類

這個數字是？

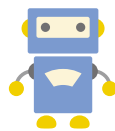
3

0? 3?

1? 2? 9?……

答案的種類有限
(離散值)

3



迴歸

明天的股價是？



?

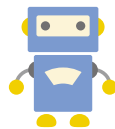
?

?

?

答案可能為任意值
(連續值)

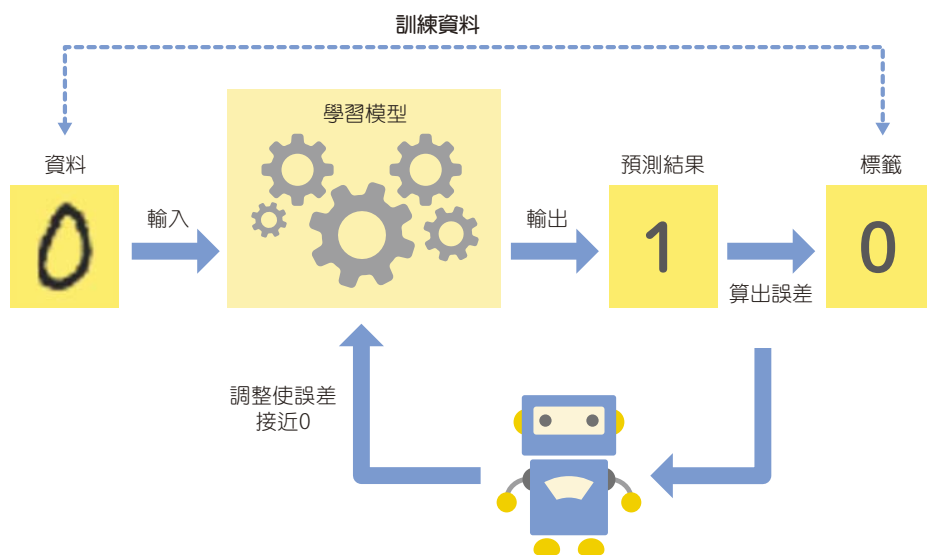
12345.6日圓



○ 監督式學習的縮小誤差

如開頭所述，監督式學習的最終目標是，正確解答未帶標籤的測試資料。為此，得先讓模型正確解答附帶標籤的訓練資料，也就是減少模型輸出的預測與標籤之間的誤差。在實際的學習中，隨著兩者的誤差接近 0，能夠正確解答的資料數會跟著增加。另外，分類通常會使用交叉熵誤差（Cross-entropy Error）；迴歸通常會使用均方誤差（Mean-square Error）。

■ 監督式學習的縮小誤差



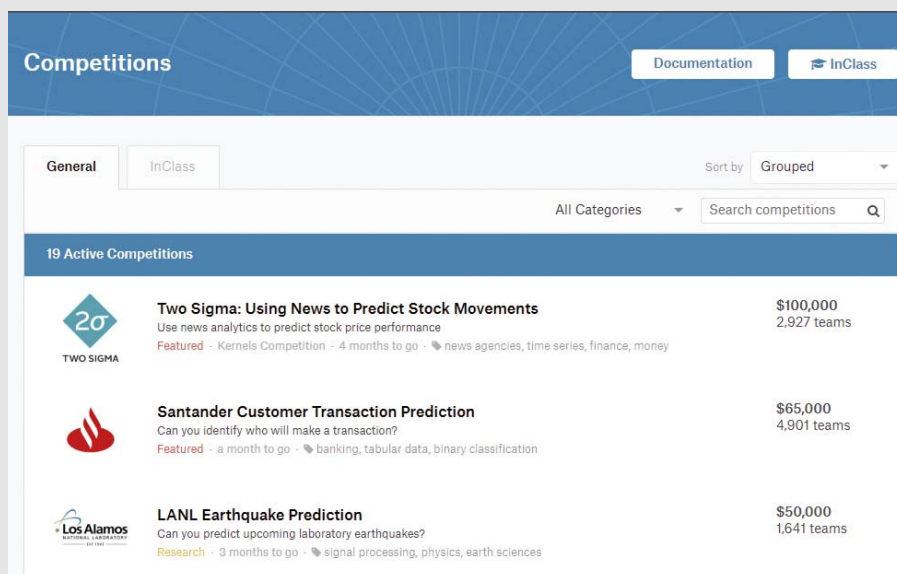
總結

- ▣ 監督式學習是指讓模型學習訓練資料的方法。
- ▣ 監督式學習的最終目標是正確解答測試資料。
- ▣ 監督式學習可分為分類與迴歸。

提升工程師技能不可欠缺的 Kaggle ？

Kaggle (<https://www.kaggle.com/>) 是一個集結全世界資料科學家、AI 工程師約 40 萬人的社群平台。特別值得注意的是，它所舉辦的針對企業或政府課題提出最佳模型，支付獎金給予最優秀模型的「競賽」活動。這項活動採用的是監督式學習，參加者能夠下載資料集。一場競賽為期 3 ~ 6 個月，排名會在獨立分數看板即時更新，直接促進參加者的動機。

除此之外，Kaggle 還有可在網頁上執程式碼的「kernel」機能，共享競賽相關的資料解析結果。然後，平台也準備了「discussion」的論壇，即便沒有參與競賽提交模型，也能夠從中獲得足夠的資訊。不過，由於 Kaggle 是英文的網站，閱讀起來可能會比較花時間。



The screenshot shows the Kaggle Competitions page. At the top, there are buttons for 'Documentation' and 'InClass'. Below that, there are tabs for 'General' and 'InClass', and a 'Sort by' dropdown menu set to 'Grouped'. A search bar is present with the text 'All Categories' and 'Search competitions'. The main content area is titled '19 Active Competitions' and lists three competitions:

Competition Name	Prize Money	Teams
Two Sigma: Using News to Predict Stock Movements Use news analytics to predict stock price performance Featured - Kernels Competition - 4 months to go - news agencies, time series, finance, money	\$100,000	2,927 teams
Santander Customer Transaction Prediction Can you identify who will make a transaction? Featured - a month to go - banking, tabular data, binary classification	\$65,000	4,901 teams
LANL Earthquake Prediction Can you predict upcoming laboratory earthquakes? Research - 3 months to go - signal processing, physics, earth sciences	\$50,000	1,641 teams

12

機器學習的基本工作程序

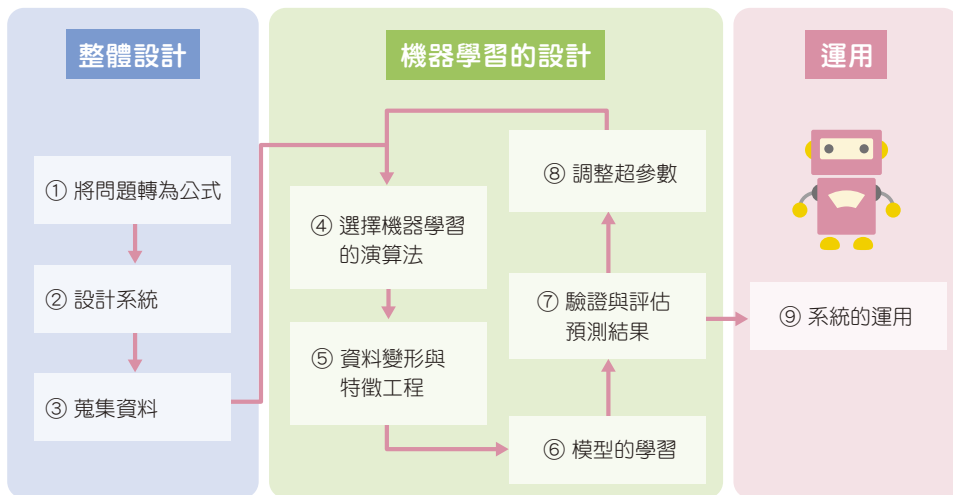
首先，我們來看系統開發的基本工作程序，並逐項解說掌握行程、認清課題等必要事項。

基本工作程序與注意要點

機器學習系統的開發跟一般的系統相比，需要較多的反覆試驗來選定驗算法、提升性能，並容易發生跨越程序間的返工（返回前一個階段重做）。因此，適當管理各程序的耗費時間非常重要。

不過，這邊的重點是，事前認清「待解決的問題適不適合利用機器學習？」。機器學習未必都會得到正確的預測，根據問題的不同，有時使用 Section04 的專家系統反而更有效率，所以必須先檢討「有沒有其他解決方案？」。

基本的工作程序



○ 整體設計

① 將問題轉為公式

會想從事機器學習的系統開發，表示內心應該抱有某種目的，比如「想要透過網路拍賣來增加收益」、「想要提升顧客滿意度」等。為了利用機器學習達成目的，必須具體深入思考「想要以機器學習獲得什麼樣的資訊」。尤其在機器學習，從輸出入的資料到演算法的選定，會因欲求事物的不同而完全不一樣，所以一開始先將問題轉為公式是非常重要的。

■ 將問題轉為公式

為了增加網路拍賣的收益……

→ 想要推薦商品

→ 預測「顧客購買某商品的可能性」 具體化到這種程度



② 設計系統

在設計系統時，需要考量機器學習細節以外的整個程序。尤其是從哪裡獲取資料、最終轉成什麼利用形式，若事前確實決定好這些事情，在後續程序中會有助於進行機器學習的反覆試驗。

③ 蒐集資料

在機器學習系統中，蒐集用來學習、預測的資料是不可欠缺的機能。除了自己當前持有的資料，還可善用政府機關、企業的公開資料或者網路蒐集的資料。關於網路蒐集資料的方法，會在 Section13 詳細講解。

23

迴歸分析

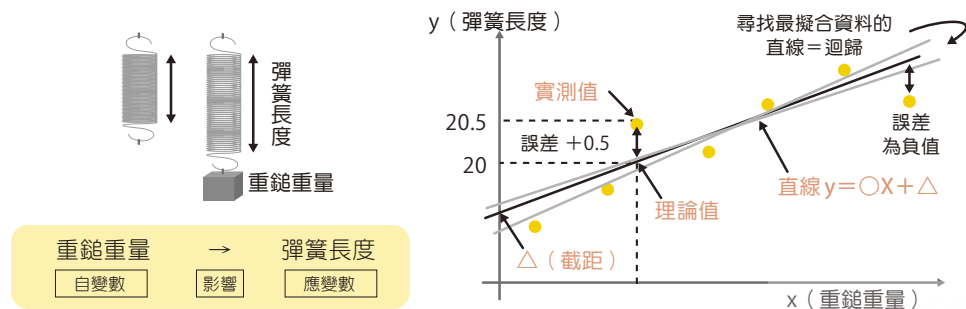
迴歸可想成是「畫出最擬合資料的直線」，下面會介紹簡單迴歸、多重迴歸、多項式迴歸、穩健迴歸等相關手法。

簡單迴歸與最小平方法

簡單迴歸 (single regression) 是以直線表達 1 個原因和 1 個結果的關係。比如，討論彈簧掛上重錘 (重量 x) 量測彈簧長度 (y) 的實驗，作為原因的重量 x 為自變數 (解釋變數)、作為結果的彈簧長度 y 為應變數 (目標變數)，在關係圖上以點 (x,y) 表示實驗結果得到的資料。數學式 $y = \bigcirc x + \triangle$ 在圖上為斜率 (係數) \bigcirc 、截距 \triangle 的直線，只要畫出最擬合資料點的直線，就能求出最佳的 \bigcirc 和 \triangle 。

「最擬合資料點」意為直線與資料的誤差總和最小。假設彈簧伸長為 20.5 公分、原長度為 20 公分，則誤差為 (y 的實測值) - (y 的理論值) = 20.5 - 20 = + 0.5。實測值為 19.8 公分的話，誤差會是一 - 0.2。然而，直接相加各個誤差的話，誤差的正負會彼此抵銷。因此，這邊會以誤差的平方和作為「誤差總和」，以便進行數學處理。如同上述，用來計算誤差總和的算式稱為誤差函數 (損失函數)，最小化誤差平方和的方法稱為最小平方法。

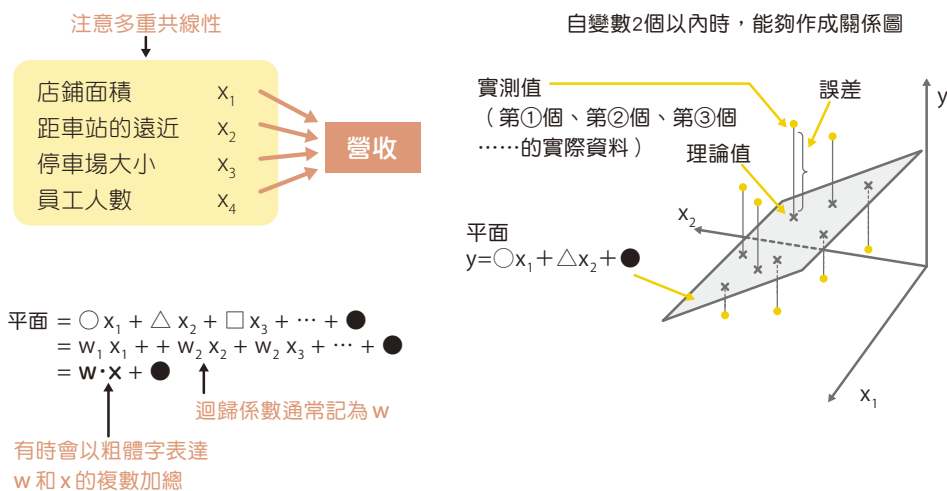
簡單迴歸



○ 多重迴歸

與簡單迴歸不同，當存在複數個原因（自變數）時，會採用多重迴歸。比如，影響零售店營收的自變數，可能有店鋪面積、與車站的距離、停車場大小、員工人數等等。假設這些自變數為 x_1 、 x_2 、 x_3 、 \dots ，則對數學式 $y = \bigcirc x_1 + \triangle x_2 + \square x_3 + \dots + \bullet$ ，尋求最佳係數 \bigcirc 、 \triangle 、 \square 、 \dots 、 \bullet 的就是多重迴歸。這條數學式在圖形上是平面，係數可想成對自變數的「影響的重要程度」——權重 (weight)，有時會直接記為 w 。此時，我們需要留意多重共線性 (multicollinearity)。比如，將降水量與降水日等兩高度相關的變數，同時當作自變數進行多重迴歸，會無法正確地迴歸分析。為了避免多重共線性，當遇到高度相關的變數組合時，必須除去其中一方。

■ 多重迴歸



多重迴歸的最小平方方法

$$\begin{aligned} \text{最小化 誤差函數} &= \text{誤差平方和} \\ &= \{y_{①} - (w_1 x_{1①} + w_2 x_{2①} + w_3 x_{3①} + \dots + \bullet)\}^2 \\ &+ \{y_{②} - (w_1 x_{1②} + w_2 x_{2②} + w_3 x_{3②} + \dots + \bullet)\}^2 \\ &+ \{y_{③} - (w_1 x_{1③} + w_2 x_{2③} + w_3 x_{3③} + \dots + \bullet)\}^2 \\ &+ \dots \\ &= \left[\text{實測值} - \text{理論值} \right]^2 = \text{誤差}^2 \end{aligned}$$