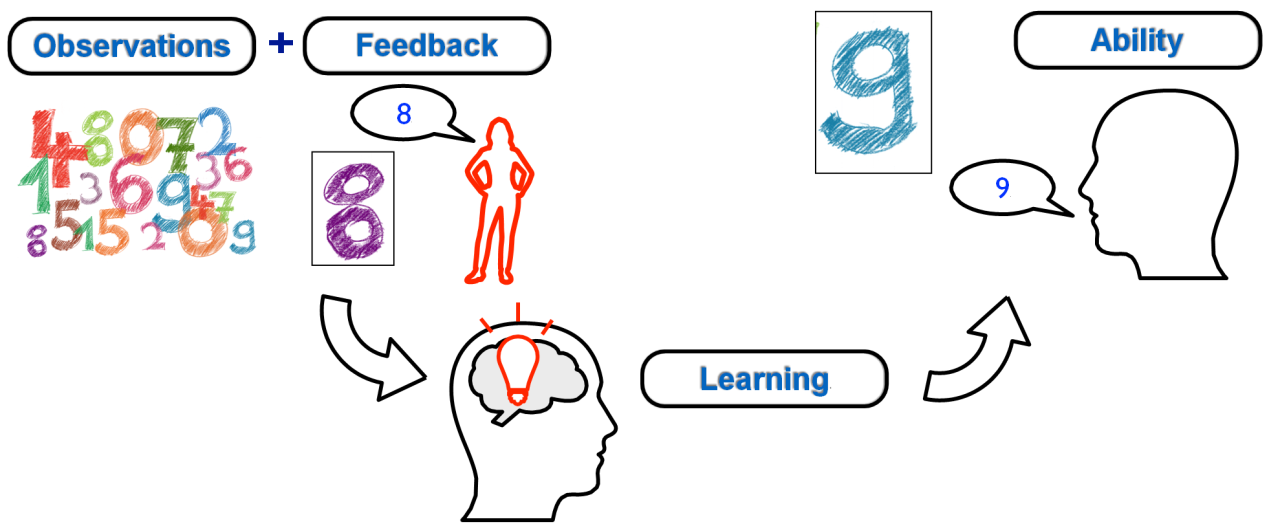
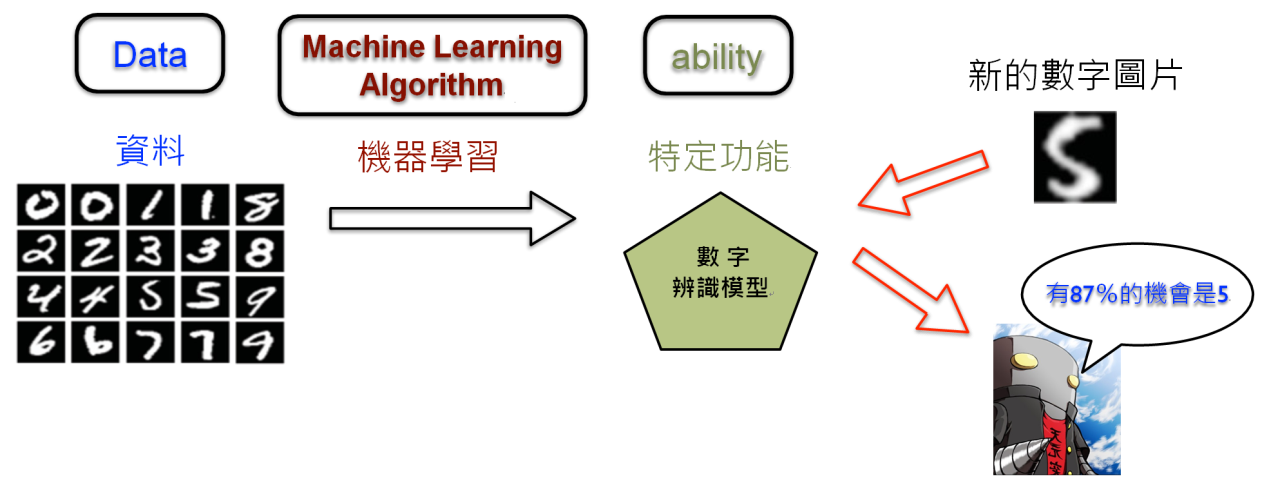
**機器學習與特徵擷取**

機器學習是人工智慧AI的一部分，著重於訓練電腦從有效的資料中學習，並根據模型演算改進，而不是按照條件式的程式碼運行作業。在機器學習中會訓練演算法尋找大型資料集的模式和關聯性，並根據該分析做出最佳決策和預測。

**從人類學習到機器學習**

⼈類學習: 藉由觀察到的事與物(observations)之經驗來學習(learning)到特定能⼒。機器學習: 藉由收集到的資料(data)來學習到特定功能

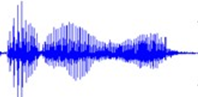


簡而言之，機器學習就是利⽤歷史資料找出⼀個函數f : X ! Y，例如:

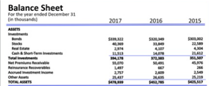
應⽤於影像分類

 => X f(X) = 貓的圖⽚(Y)

應⽤於語音辨識

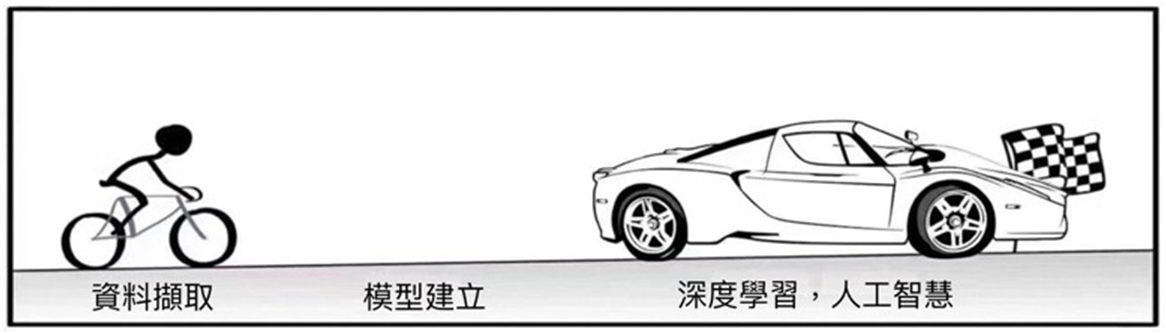
 => X f(X) = 我是貓(Y)

應⽤於企業核貸

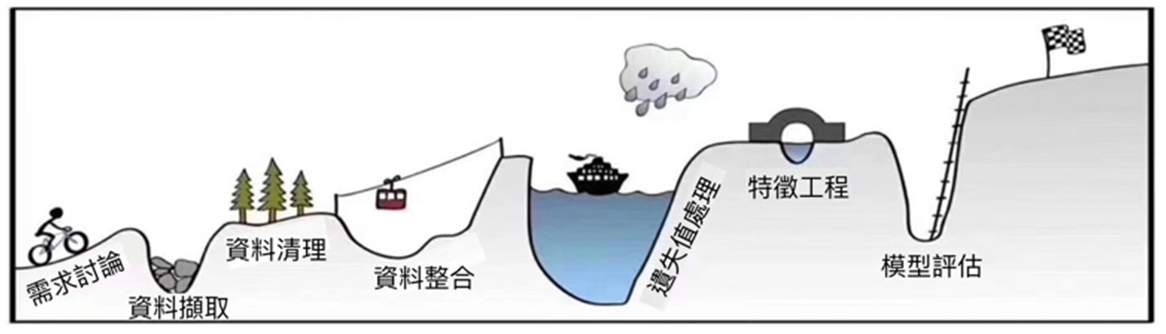
 => X f(X) = 優質貸款客戶(Y)

**機器學習運作概念**

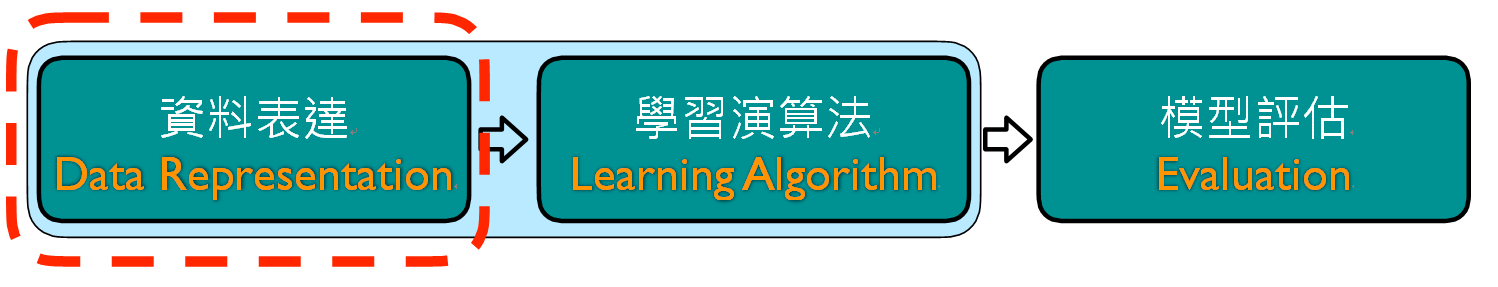
大眾的見解



實際的過程



**資料表達**



⼀般來說，資料可分成兩個部分：

特徵 (Feature)：⽤來描述每⼀筆資料，通常會⽤ X 來表示

標記 (label)：⽤來表示每⼀筆資料所對應的輸出，這個輸出樣式可以有不同的狀態(可能是類別或者實數值等)，通常會⽤Y 來表示。



Y: 男 or 女

X:體重

身高

髮色

髮長

膚色

鬍鬚

臉型……

Data

• 使用這些特徵來描述數據實例。

• 可以是離散的、數字的等。

Data



如果Y:是否為盜刷?或是Y:未來是否為在百貨類消費?

試問要如何從Data中取出適當的特徵X呢?

在資料表階段。如何提取好的特徵是機器學習中的一項重要任務。

**特徵擷取**

例如: 信⽤卡刷卡紀錄



Y: 未來是否為在百貨類消費？

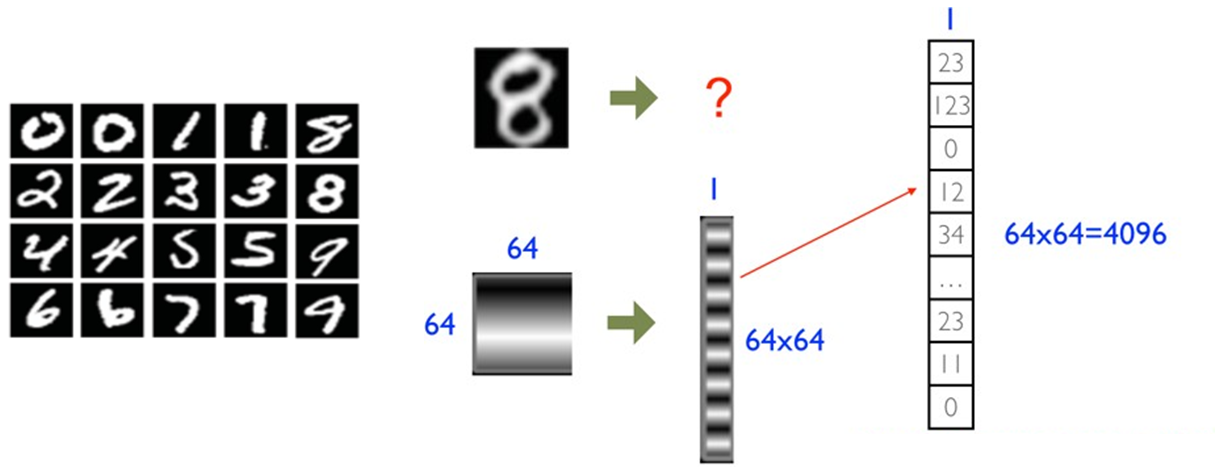
X: 過去⼀個⽉在百貨類累積消費⾦額

過去⼀個⽉在百貨類累積消費次數

過去⼀個⽉信用卡總消費⾦額

例如:影像資料

最簡單的方法是直接使用像素值，人臉識別和手寫數字識別中的通用方式。



* 過