第二週：

第二週課程正式進入了機器學習的領域，也大概講述了一個機器學習的流程和目的、可能遇到問題等。

1. 什麼是機器學習？

時常有許多人將人工智慧和機器學習混為一談。但機器學習和人工智慧的關係其實應該如圖所示：

可以發現，機器學習只是人工智慧的一部份。因此雖然機器學習對於建構一個人工智慧來說相當重要，但卻萬萬不能將其和人工智慧畫上等號。

2. 機器學習的類型

機器學習，依照問題的不同和學習方式的不同，大約可分為三種：

① 監督式學習(Supervised learning)

定義：將所有資料標註標籤(相當於正確答案)，以作為機器的參考。標籤多為人為定義，是較基礎的訓練方式，但在定義標籤時會較繁複。

目的：學習資料和標籤的關係

舉例：學習房屋坪數和標籤的關係(線性回歸問題)

② 非監督式學習(Unsupervised learning)

定義：所有的資料都沒有標籤，機器需自行依照資料特徵進行分類。雖然

不須人為上標籤，對人類是輕鬆的，但對於機器的訓練十分困難。

目的：學習資料間的隱含的結構

舉例：社交網路的關係(分群)

③ 增強式學習(Reinforcement learning)

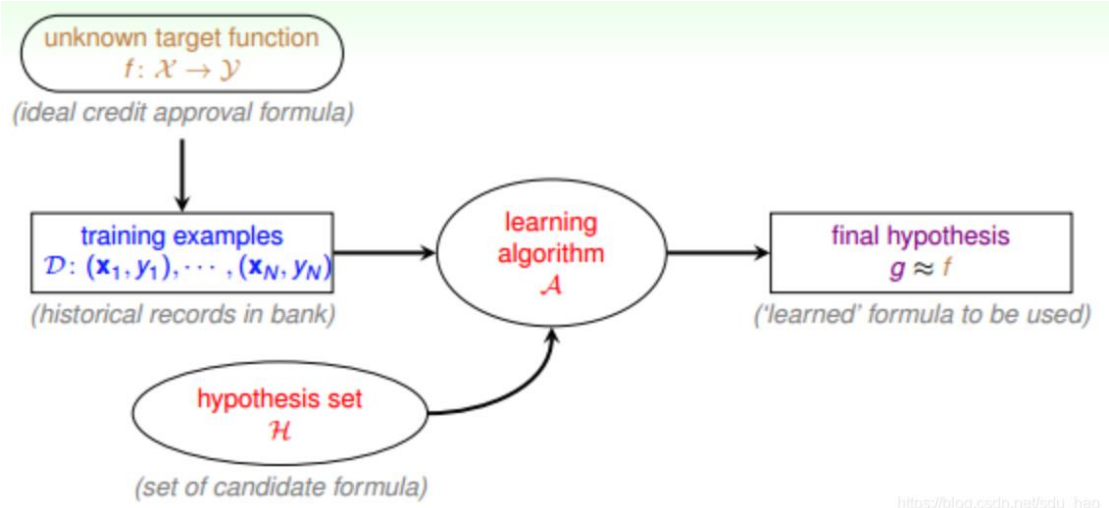
定義：機器透過與環境的互動來學習如何得到最大的效益。用以訓練的資料並無標籤，但對於每一次的行動，將給予回饋和獎勵，機器會這一系列回饋來修正自己的行動，並且得到正確結果。

目的：學習如何選擇動作已得到最大效益

舉例：Open Ai pong(兵乓球遊戲)

3. 機器學習的流程

機器學習的大致流程如下圖：



① 目標：機器學習的目標，是找出一個和真實的目標函數 $f(x)$ 相近的函數 $g(x)$，並用此函數來預測輸入資料的結果。

① 訓練資料：

由真實模型產生資料，以作為往後訓練機器的樣本。

② 假設集合(hypothesis set)：

在定義學習任務後，選擇適合該任務的學習模型。每個模型皆有對應的假設集合，而假設集合中又包含了多個假設，我們挑選一個或多個假設，以和訓練資料一起放入學習演算法中訓練。

舉例：線性模型、神經網路皆是屬於假設集合。

④學習演算法

有許多不同種類的學習演算法以應對不同的問題。學習演算法的目的是使用訓練資料和挑選出的假設找出和的真實模型最小誤差。

Δ得出的函數 $g(x)$，就算在訓練資料上表現好，也不一定在所有資料上都表現得好，因此，訓練資料要避免噪音(noise)和偏誤(bias)。

4. 線性模型與特徵轉換

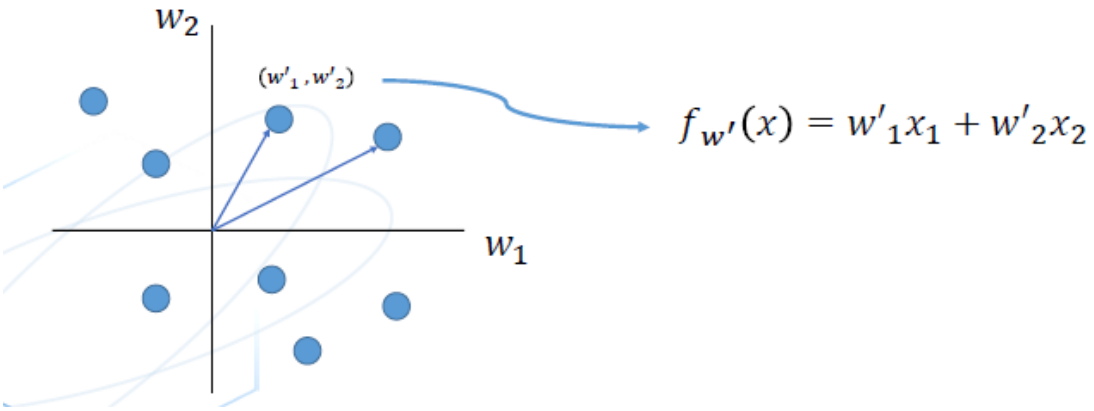
① 線性函數

線性函數：$f(x) = W_1X_1+W_2X_2+\cdots\cdots+W_nX_n = W^TX$

此 $f(x)$形成一個線性假設空間 →$\mathbb{W}$ 的值不同，可形成不同的 $f(x)$。

W^T 包含了從 W_1~W_n的所有向量。

線性假設空間由許多的向量構成，並可由座標表示：



② 特徵轉換($\phi(x)$, *kernel function*)

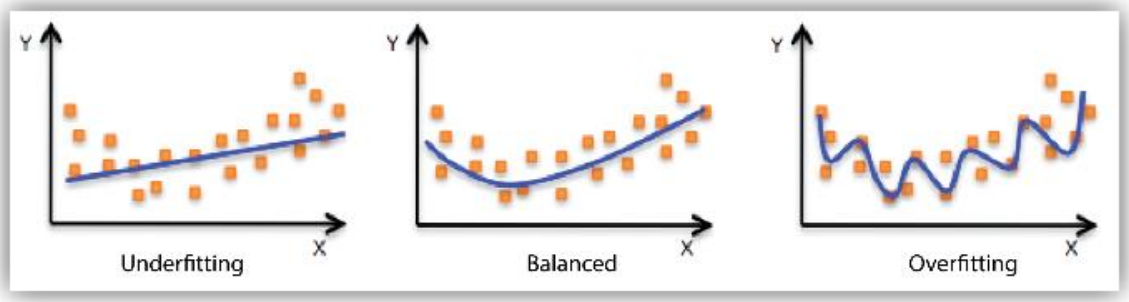
用途：

1. 將非線性的問題線性化，使其有辦法以線性函數表示。
2. 將資料轉換到新的空間，再以線性模型處理。
3. 經過特徵轉換，可以減少時間複雜度

5. Overfitting(過擬和)

機器學習的最終目的，便是達到 generization，意即能符合大部分情形的資料。但有時訓練結果只會和訓練資料相符，和其他資料卻不能達到好效果。這種情形稱之為 overfitting。

我們以圖片比較 overfitting、正常情況和 underfitting：



1.Overfitting 的成因和特徵：

在上圖最右邊的圖表中，我們可以看到 overfitting 的形式：呈現出的圖形和訓練資料幾乎完全符合，因此呈現的是變異大且有噪音(noise)的情形。這時假如加入新的資料，可能結果就會完全不同。因模型受訓練資料的影響太大，導致無法在其他資料上表現好。

①Overfitting 的解決方式：

為了避免 overfitting，我們需要做的就是預先處理資料和減少模型的複雜度。若想減少模型複雜度，可以使用 Regularization(正則化)：此方法可以限縮假設空間(模型)可以挑選的範圍，以達到減少計算結果太複雜的目的。

2.Underfitting

上圖最左邊的圖表中，呈現的擬和曲線和資料間的關係相當不明顯，無法準確呈現資料間的關係，這種情形稱為 Underfitting。會造成這種情形的原因可能是模型的參數過少，或資料結構過於簡單。想要解決這種情形，便是增加模型參數(激活函數便是類似用途)或增加資料結構的複雜度(例如變更輸入的特徵)。

參考資料：

1. 林軒田教授機器學習基石 Machine Learning Foundations 第 8 講學習筆記
https://blog.fukuball.com/lin-xuan-tian-jiao-shou-ji-qi-xue-xi-ji-shi-machine-learning-foundations-di-ba-jiang-xue-xi-bi-ji/

2. Quora
https://www.quora.com/What-are-the-key-trade-offs-between-overfitting-and-underfitting