

intro

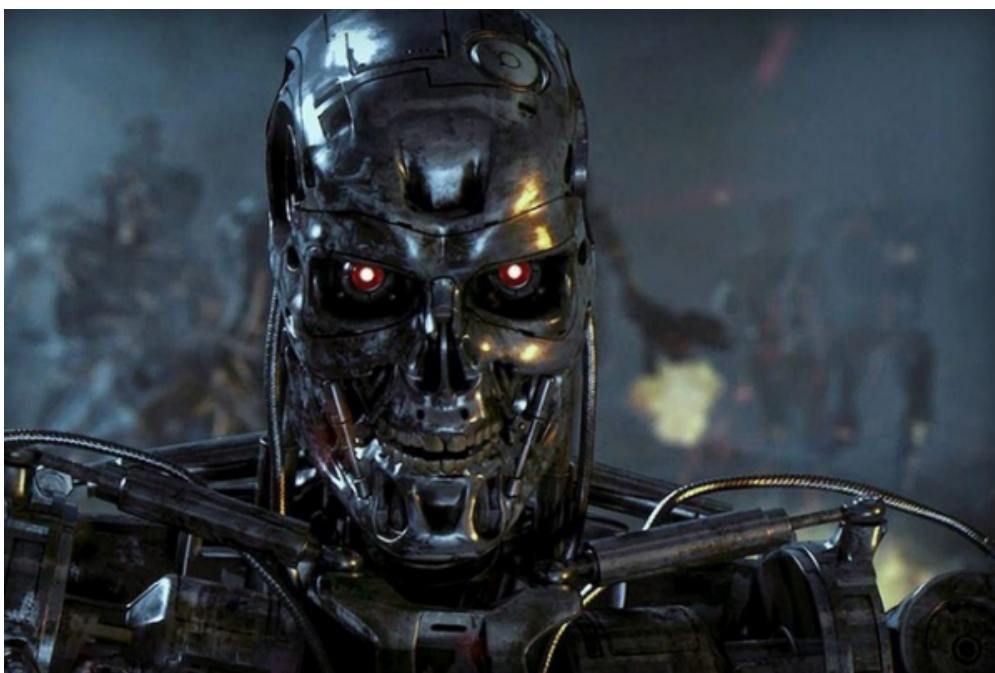
2018 年 6 月 22 日

Machine Learning

1 機器學習介紹

說到機器學習，我們就得先定義幾個名詞

1.1 AI (Artificial Intelligence)



機器是否能擁有智慧呢？

如果可以有的話，要怎麼定義擁有智慧這件事呢？

著名的電腦科學家艾倫·圖靈提出一個很有名的測試方法，我們直到現在都還在用它，他叫做圖靈測試。

1.1.1 圖靈測試

描述：如果一個人（代號 C）使用測試對象皆理解的語言去詢問兩個他不能看見的對象任意一串問題。對象為：一個是正常思維的人（代號 B）、一個是機器（代號 A）。如果經過若干詢問以後，C 不能得出實質的區別來分辨 A 與 B 的不同，則此機器 A 通過圖靈測試。

白話文就是：你分出來這台機器跟人類有什麼不同！

從這測試被提出後，我們就開始用圖靈測試當作 AI 的最高守則

但是我們最後發現，一個機器要在所有領域通過圖靈測試，也就是跟真正的人類一樣，以現在技術要達成是做不到的！！

所以我們把通過圖靈測試的機器分成兩種：1. 強人工智慧 2. 弱人工智慧

1.1.2 強人工智慧

如果今天我製造出一個機器，他可以在所有領域都通過圖靈測試，透過學習學會任何判斷，我們就稱為強人工智慧。

但大家普遍的認知，在現今的科技是無法做到的

1.1.3 弱人工智慧

今天有一個機器也可以通過圖靈測試，但只限於某些領域，可能是語言分析領域或者是影像分析領域，那我們就稱它為弱人工智慧。

這也是我們今天在討論的人工智慧！

1.2 AI 發展史

1.2.1 專家系統

有一陣子，大家嘗試著把所有知識化成一條一條的規則輸入到電腦裡，但後來大家發現兩個難處

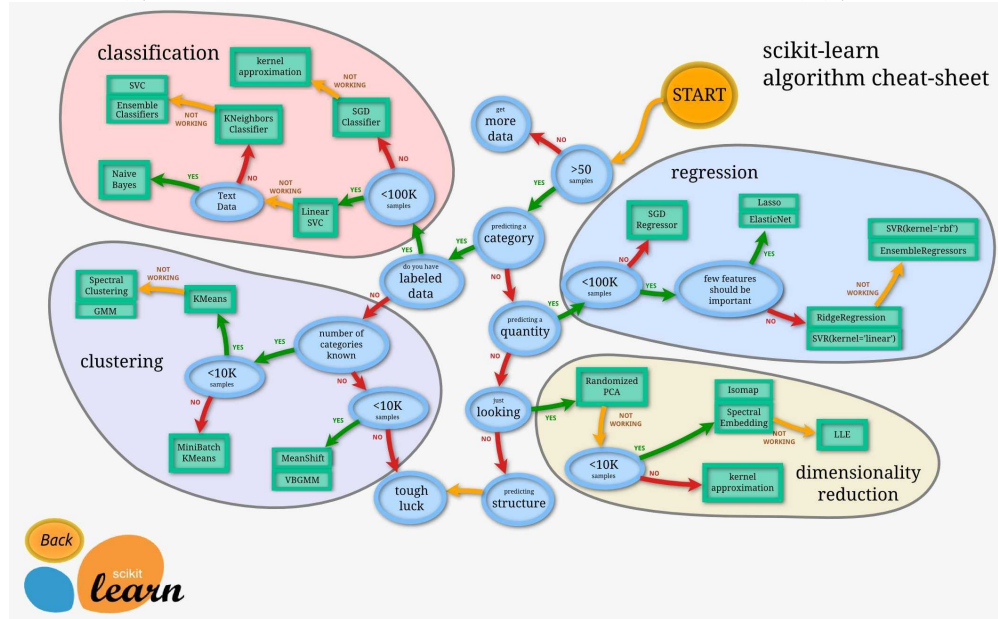
1. 人類的知識量太大太大了
2. 有些抽象的知識 (ex. 飛的概念) 根本不知道怎麼化成一個描述

所以專家系統就沈寂了，同時起來的就是透過機率和統計模型來試圖根據得到的資料做出判斷，也就是我們的機器學習

1.2.2 機器學習

簡單來說，就是機率與統計，透過擁有的資料對未知的資料做出判斷，非常多的演算法被提出，每個演算法都有自己擅長的領域以及適合的資料

有哪些演算法呢？給各位一些概念，後續我們會就裡面經典的演算法介紹



1.2.3 深度學習

在機器學習的發展史我們最後遇到一個困境，我們對於預測抽象的概念，譬如什麼是一個椅子，或者什麼是信仰，雖然可以預測

但是預測的結果不盡理想，而且做出來的模型解釋性很差，所以就有人想了一個概念，我們何不模擬人類神經運作的方式呢？

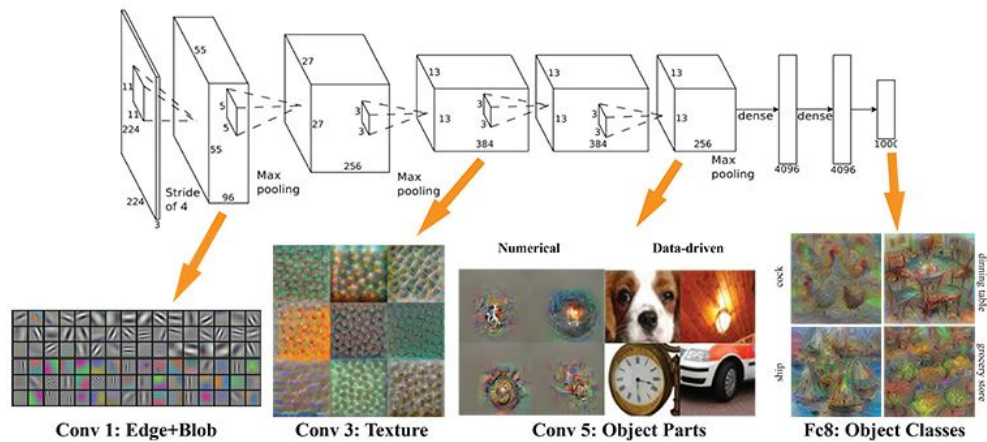
雖然神經運作的方式至今沒人理解，但大家有個共同的共識，就是人類在學習的過程中是由具象 (椅子的邊角) 學習到抽象 (什麼是椅子)

所以機器學習其中的一種演算法 (人工神經網路) 就此興起，因為這種演算法可以在中間加入很多層的過濾，看起來就好像我們的神經系統

一層一層從具象到抽象!!

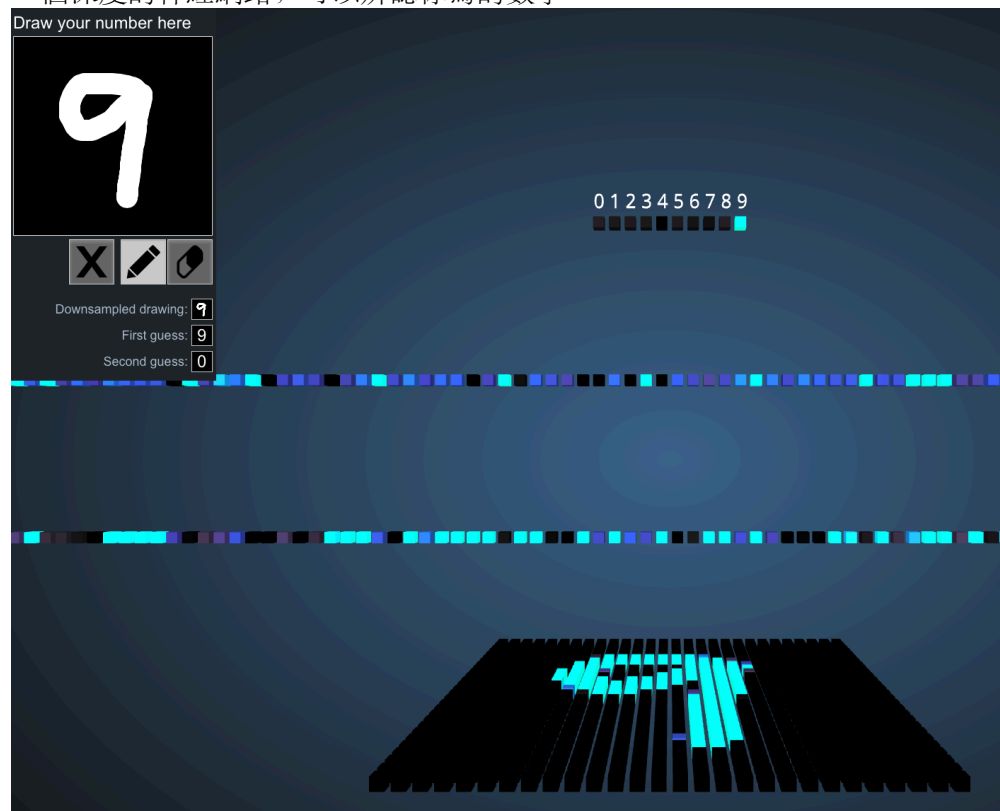
所以其實所有的人工神經網路都可以擴展成深度的人工神經網路

給大家一個例子來看看什麼是具象到抽象，圖擷取自 <http://xilinx.eetrend.com/article/10827>



有興趣的讀者可以玩玩看 <http://scs.ryerson.ca/~aharley/vis/fc/>

一個深度的神經網路，可以辨認你寫的數字



1.3 學習類型

1.3.1 資料類型

根據我們手上的資料類型，可以分成三種處理方式

1. 監督式學習：資料已經被填上了正確答案，例如：每一首詩已經被註明是哪個詩人寫的

2. 非監督式學習：有時候因為資料太多，或者沒有一個可以填入正確答案的手段，我們就只能靠電腦自己去填已有資料的答案。
3. 強化學習：算是介於中間，是有正確答案的，不過這正確答案是從環境產生，例如：教電腦玩遊戲，遊戲中的分數其實就是你每一個資料的答案

前兩種方式是接下來我們討論的重點，當然你已經可以預想到，有答案的資料總是比沒有答案的資料訓練效果要好

就像老師已經給了你十題例題的答案，你做未知的第十一題就比較容易對!!!

1.3.2 預測答案類型

根據我們要預測的答案，又可以分成兩種預測

1. 分類：預測的答案不是連續性的數字，而只是單純類型預測 (ex. 預測寫詩的詩人)
2. 迴歸：預測的答案是連續的數字 (ex. 預測房價跟附近環境的關係)

1.3.3 應用

接下來我們看看資料類型 + 預測答案類型對應哪些我們現實生活的應用！

