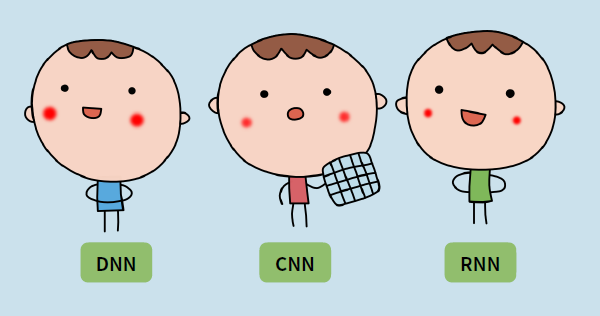
**可記憶的深度學習RNN懶人包**

深度學習應用除了圖形辨識，預測分析外，還有一個進步的發展-語意辨識，關鍵的技術為可記憶的學習模型，機器使用 CNN 來消化處理影像，相當於眼睛的角色，以辨識不同物體；而 RNN 是數學計算引擎，相當於耳朵和嘴巴的角色，以解析各種語言模，本文我們就來聊聊RNN的原理及常見應用。



PS:DNN是ANN更多層連接的模型。

**循環神經網路Recurrent Neural Networks（RNN）**

RNN是一種專門設計來解決時間相關問題的神經網路。什麼是時間相關問題？

就是當下的答案會受過去的答案影響，而且也會影響未來的答案的一種特殊情況。也可以說是次序相關問題。

舉例來說：  
如果中午訂便當的流程如下：

星期一:麥當勞、星期二:肯德基、星期三:胖老爹、星期四:摩斯漢堡、星期五:漢堡王。

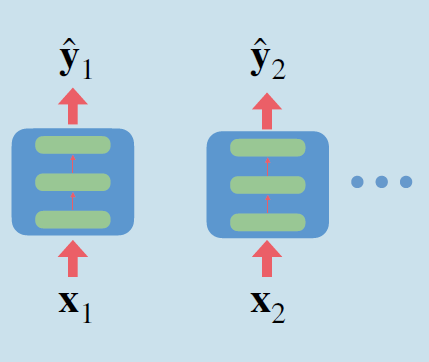
就算忘記今天星期幾，只要記得昨天吃了什麼就可以推斷出今天要吃什麼。那就會說這個答案是有時間相關性的。

而只要答案有所謂的時間相關性，那他就是一個時間相關問題，可以用RNN來解決。

我們用下列簡要機制說明RNN是如何運作的。

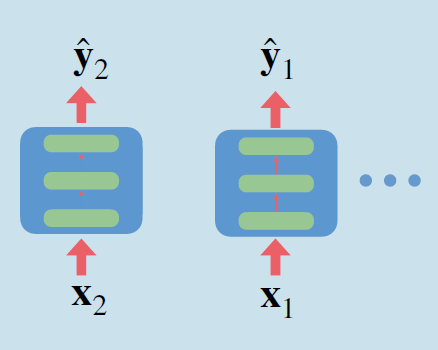


⼀般的神經網路⼀筆輸⼊和下⼀筆是沒有關係的。



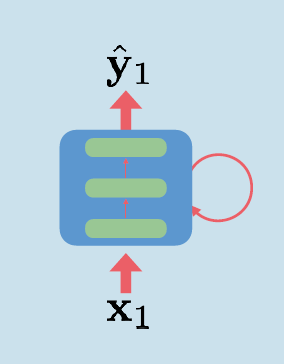


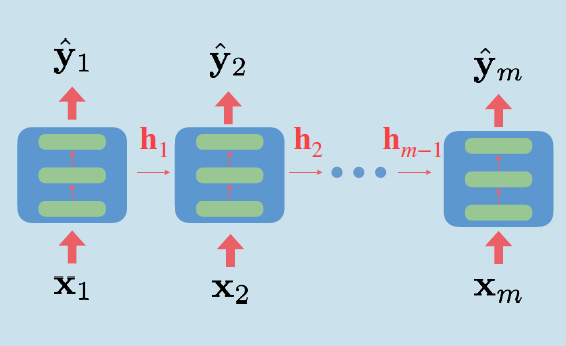
也就是說，如果把輸入順序調換，其結果是一樣的。





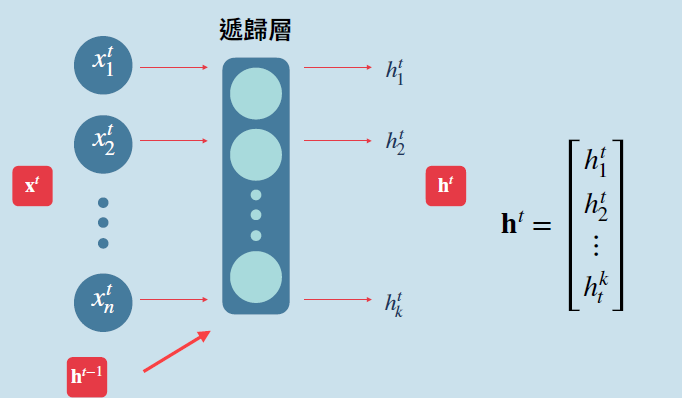
但是RNN會把上一次的輸出也當作這一次的輸入。我們稱之為Hidden States。

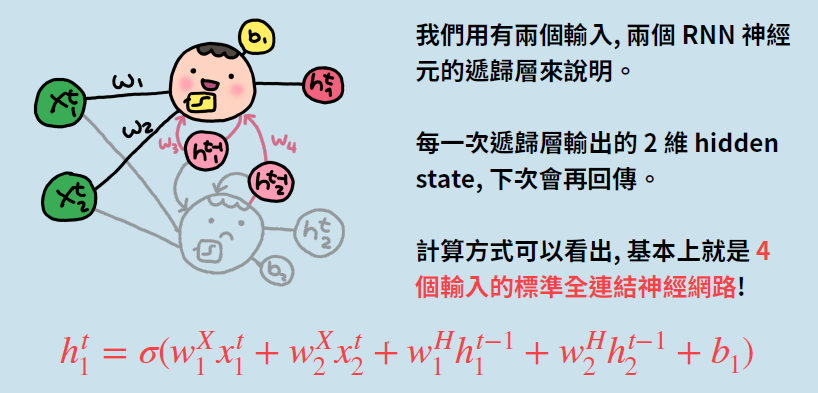






這個 hidden state 向量下次會當成輸⼊的⼀部!

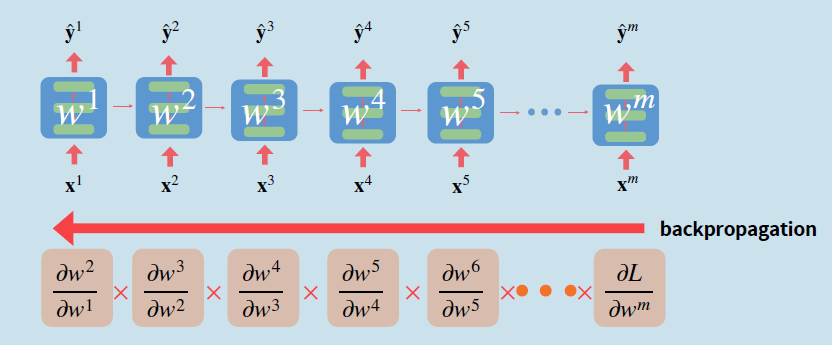






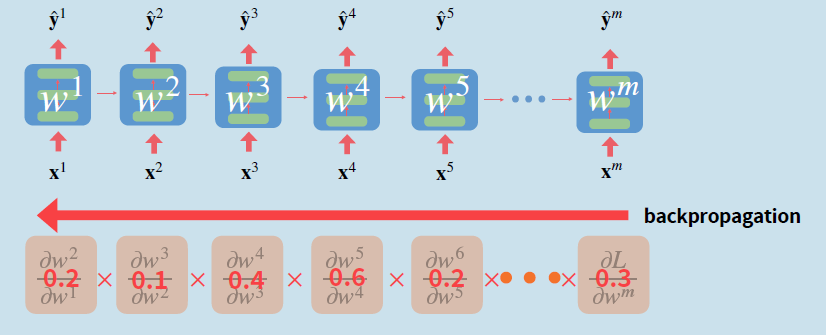
RNN雖然可以記憶之前的歷史訊息，但也因此有一個非常致命的缺點，很容易梯度消失!我們來見說明原因及改善方法。

**RNN的訓練BPTT及LSTM/GRU**





訓練的時間我們先把整個過程看成⼀個很深的神經網路, 把每個時間點的 看成不同的參數去調整。每⼀個參數要調多少再平均, 就得到要調整的⼤⼩。





在做backpropagation時，還沒乘到最前面，往往就變成0了，這個問題就叫做梯度消失(Vanishing gradient)。

解決訓練問題，可用兩種方式:

LSTM ( Long Short Term Memory )

GRU ( Gated Recurrent Unit )

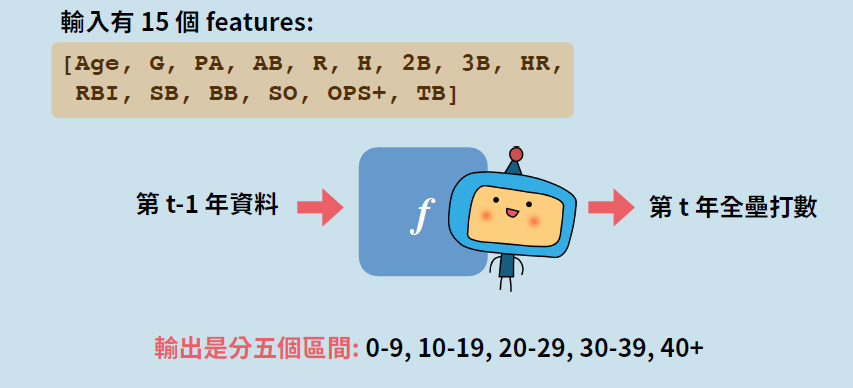
讓Hidden State與歷史資訊間間呈現線性而非乘積關係，避免梯度消失。

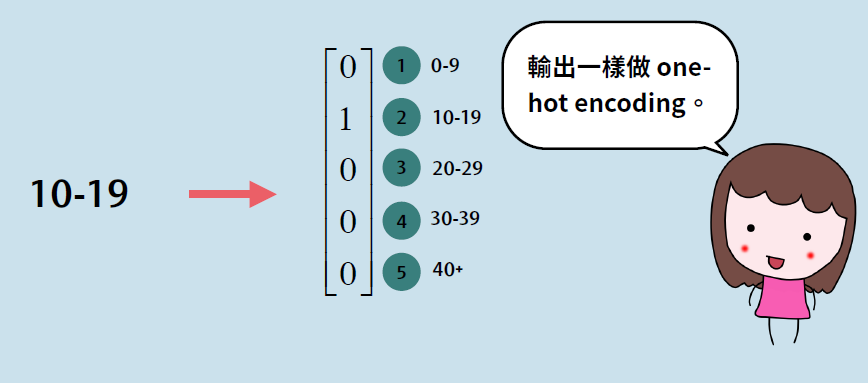


我們用一個例子來說明前述的立論。



我想知道某位 MLB選⼿新的球季可以打幾隻全壘打?





使用Tensor Flow神經網路函數學習機的設計:

1.輸入 15 維向量, 輸出 5 維向量 (區間)。

2.一層LSTM 隱藏層。

3.輸出層做 softmax(原始的RNN, 我們習慣叫它 Vanilla RNN, 在 Tensor Flow 中是SimpleRNN)。

4.訓練時每次用 10 年資料。





**總結:**

RNN最重要的應⽤之⼀， ⼤概就是可以⽤⾃然語⾔交談的對話機器⼈。

其實對話機械人模式可以做更多…..

輸⼊不⼀定要⽂字, 是影⽚ (⼀張⼀張的圖) 也是可以 的! 輸出還是可以為⽂字, 最常⾒的⼤概是讓電腦說 影⽚中發⽣什麼事(Video Captioning)。

或是:

自動翻譯

生成文章等等。



* 過