

神秘的典禮

壹、前言

這一天，穿著一襲禮服的 Win10 參加了 Microsoft 學校的校慶典禮，當大家高興地聊著天吃著東西時，突然一個愚笨的少年衝了過來，用兇狠的眼神瞪著她，他自稱他原本是 Microsoft 學校裡的一個工友，只是因為他太愛國了，屢屢遭到學生投訴，今天他來就是要來叫 Win10 大美女幫他討回公道。

他說：「反正同學們對我的成見已深，我即使回到了學校也是不太可能好好工作的，妳如果沒辦法幫助我找到第二份工作，我就會剝奪你最珍惜的東西，然後這間學校也會付之一炬。」這時，所有在典禮的同學們呆掉了，這是他們第一次這麼靠近死亡，這麼希望能夠離開這間學校，大家的性命都託付在 Win10 的手裡，Win10 緊張的說不出話來，那個少年說話了：「不要再拖了！時間一直在流逝，妳快快想辦法，我只給妳一個禮拜的時間，如果沒幫助我找到第二份工作，後果不堪設想！」Win10 靠著她的智慧與勇氣，想出了一個方法，「以人工智慧協助**求職者面試預測**」，因此她號召了全校最優秀且頂尖的科學、資訊人才，開始著手訓練這個模型。

經過她們不眠不休的測試與訓練，事情似乎露出了一絲曙光，Win10 親自教導同學們如何計算矩陣及向量，同學們一聽了她的講解，恍然大悟，也讓整個模型開發越來越順利，但，她們卡關了，他們卡在一個十分重要的點”激勵函數”，沒有了激勵函數，深度學習僅僅是數個矩陣進行運算，只有經過非線性的激勵函數，才能真正讓模型有意義，同樣也是資訊能力超強的你們，加入她們的行列吧！幫助思維謹慎、具有領導能力的 Win10 守住校園吧，校園是否會被這個愚笨的不良少年摧毀，全靠你們的智慧了。

類神經網路中都需要一個激勵函數(Activation Function)，激勵函數的主要用途是為了解決輸入和輸出的線性關係、並增加模型的表現力，因此有許多不同的激勵函數，這些激勵函數各自有各自的優缺點，而其中以 ReLU 函數成為現代深度學習的主流，其他的函數有許多問題，像是 Sigmoid 和 Tanh 函數時常會發生梯度消失的問題，再者，ReLU 相較其他函數而言，十分容易計算，也可以省下許多資源。

深度學習時常用到的目標函數(Object Function)都是使用損失函數(Loss Function)，也就是下面題目要叫你算的 Loss，一個模型的好壞通常與損失函數的設計也有很大的關係，回歸函數被分類為兩個部分，回歸及分類，對於回歸問題，Loss Function 常用的是 MSE 均方誤差、MAE 平均絕對值誤差這兩種，像是第 59 屆科展電腦與資訊學科的第一名「**以類神經網路為輔助自動生成小提琴演奏骨架**」就是使用 MSE 作為 Loss Function，而在分類問題上，選擇的通常就是 cross-entropy，當然，與激活函數一樣，每一個函數都有自己的優缺點，在開發時就是要對於模型多測試，找出最適合模型的函數。

參考下圖 Image1 可以發現常見的三種激活函數的圖形。此題只需用到第三個 ReLU 函數，我們可以推導出他的一般式：

$$ReLU(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

OR

$$ReLU(x) = \max(x, 0)$$

再來，請參考 Image2 的例題圖片及 Image3 的矩陣乘法範例圖片，本題由於不會使用到陣列，因此會固定向量、矩陣的大小，且你只需要幫忙求出 Loss 函數（這裡使用的是 MSE 均方誤差）的值即可，不需要進行偏微分（我真的人超級好的，請感謝 Win10）。

附圖：

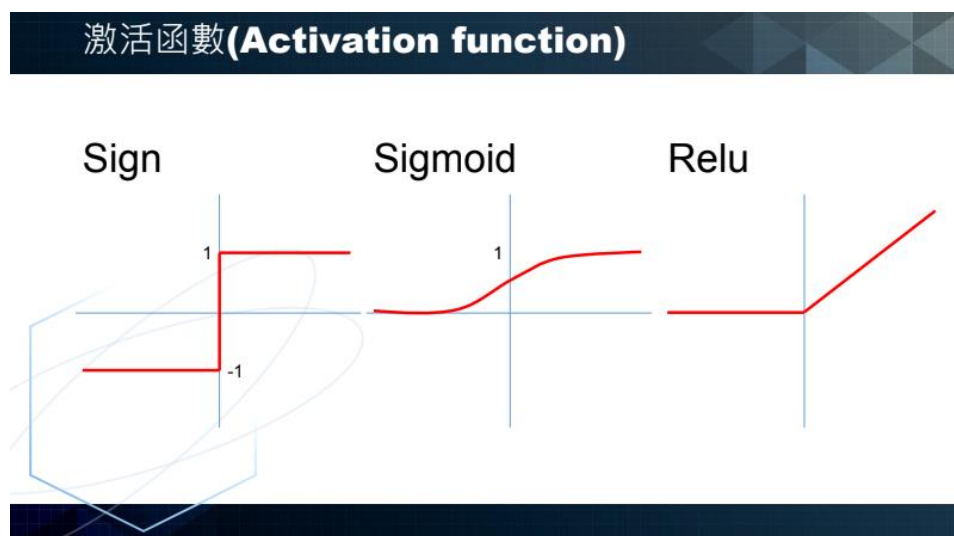


Image1 、 本圖出自：交大人才培育計畫 Week4 講義

具有一層隱藏層的神經網路，假設輸入層有 3 個節點，輸入數據 X 中有 3 筆數據，其標籤為 y ，隱藏層有 2 個節點，隱藏層權重矩陣為 W_1 ，線性組合 $Z = XW_1$ ，經過激活函數 Relu 的值令為 $L = \text{Relu}(Z)$ ，輸出層有 1 個節點，的權重矩陣為 W_2 ，線性輸出

$O = LW_2$ ，將輸出值與標籤計算 MSE loss($loss_{MSE} = \frac{1}{2} * (O - y)^2$)，令回傳損失 $G_2 =$

$$\frac{\partial loss_{MSE}}{\partial W_2} \text{ 與 } G_1 = \frac{\partial loss_{MSE}}{\partial W_1},$$

$$X = \begin{bmatrix} 2 & 1 & 0 \\ 3 & -2 & 1 \\ 2 & 0 & 1 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, W_1 = \begin{bmatrix} -1 & 0 \\ 2 & 1 \\ -2 & -1 \end{bmatrix}, W_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Image2 、 本圖出自：交大人才培育計畫 期末測驗

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} (aA+bB+cC) \\ (aD+bE+cF) \end{bmatrix}$$

Image3 、 矩陣乘法的運算

貳、輸入說明

本題假設為有一隱藏層的神經網路，輸入層有 3 個節點，輸入數據 X (一) 有 3 筆數據，標籤為 Y (二)，隱藏層有 2 個節點，隱藏層權重矩陣為 W_1 (三)，線性組合 $Z = XW_1$ ，經過激活函數 ReLU ，令 $L = \text{ReLU}(Z)$ ，輸出層有一個節點，其權重矩陣為 W_2 (四)，線性輸出 $O = LW_2$ ，計算 $\text{Loss} =$

$$\frac{1}{2} \sum_{i=1}^n (O_i - Y_i)^2。$$

(題目與上面附圖大致相同)

一、第一行會輸入 9 個數字，為矩陣 X ，格式請參考下面解說。

輸入 a b c d e f g h i

其意涵為：

$$X = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

二、第二行會輸入 3 個數字，為矩陣 Y ，格式請參考下面解說。

輸入 j k l

其意涵為：

$$Y = \begin{bmatrix} j \\ k \\ l \end{bmatrix}$$

三、第三行會輸入 6 個數字，為矩陣 W_1 ，格式請參考下面解說。

輸入 m n o p q r

其意涵為：

$$W_1 = \begin{bmatrix} m & n \\ o & p \\ q & r \end{bmatrix}$$

四、第四行會輸入 2 個數字，為矩陣 W_2 ，格式請參考下面解說。

輸入 s t

其意涵為：

$$W_2 = \begin{bmatrix} s \\ t \end{bmatrix}$$

參、輸出說明

計算並輸出 Loss 函數的值。

輸出到小數點後第二位。

行尾換行

肆、範例測資

輸入：

2 1 0 3 -2 1 2 0 1

1 0 -1

-1 0 2 1 -2 -1

1 -1

輸出：

22