

- 一、 Sigmoid 他會主要會把 $(-\infty, \infty)$ 的 input，變成 $(0,1)$ 的 output，且各個 input 會被個別處理，所以結果會相互獨立。

Softmax 的輸出值相互關連，且會把多個 class 的 output 轉成一個機率分布，使他機率和是 1。

當我們選擇 binary classification 的時候，會選擇 sigmoid，因為他可以幫我們個別處理每一個 input，而不會與其他 input 有所牽連。

當我們選擇 multi-class classification 的時候，會選擇 softmax，因為他可以幫我們在所有的 class 中，選擇一個機率最高的 class。

- 二、 learning rate 太高、minibatch 的 size 太小、沒有 regularization

- 三、 大的 learning rate，會讓 model 可以在每個 step 有更大幅度的更新，然後有機會減少所需步數，所以可以減少時間，但是如果設太大可能會直接忽略最優解的部分；小的 learning rate，則相反，所以可能會需要更多的時間去達到最優解，但可以更穩地找到最優解。

大的 batch size 可以讓更多資料同時被處理，所以可以讓資料的速度更加快速，但是他需要更多記憶體。

小的 batch size 需要更多的 iteration 去執行整個 dataset，會增加整體的時間，但是可以減少所需的記憶體。

- 四、

