1. Sigmoid他會主要會把(-inf, inf)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muti-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。

1. learning rate太高、minibatch的size太小、沒有regularization
2. 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、

一張含有 文字, 圖表, 行, 螢幕擷取畫面 的圖片

自動產生的描述