

## 1. 遇到的困難與解法

一開始花了許多時間在理解 `template` 的架構以及 `class Dense, Activation` 之間的互動關係。

Basic part 在寫 `Activation-forward-softmax` 的時候參考網路上的做法，因為網路上跟作業的 `data representation` 不一樣導致我一時之間找不到問題，後來將 `np.sum` 對 `column` 加總就順利解決了。

Advance part 一直遇到 `nan`，原本以為是哪邊的 `forward` 錯誤或是 `softmax` 算錯，結果最後發現是 `y_train` 在 `preprocessing` 沒做好，改成 `one-hot` 跟正確的 `shape` 後就解決了。

## 2. structure of classifier

Binary:

`30 * 30 * 30 * 1` 的 network

activation functions: `relu, relu, sigmoid`

5000 iterations

learning rate = 0.1

Advance:

`784 * 10 * 10` 的 network

activation functions: `relu, softmax`

1000 iterations

learning rate = 0.05

batch size = 64

## 3. efforts on tuning parameters

不管怎麼調整 `num of iterations` 或是 `hidden layer` 的 `node` 數，我發現 `cost` 到最後都會趨近於一個值就難以再往下了；也發現 2 個 `hidden layer` 的 NN 表現比 1 個還要好，但更多層的 `hidden layer` 表現不一定比較好。