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。

1. 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

1. 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。