

# Hw3 bonus tensorflow

B05902109 資工三 柯上優

## a. 做了那些比較的實驗

對於不同的實驗，我以下將以改動的各项參數一起解釋。

### - Random seed and pretrain model:

為了能重現實驗結果，我將 **random seed** 設為一任意數字 **7122**。而既然 HTC 方已經有 **pretrain model**，我就心懷感激地使用了。

### - Optimizer:

粗淺嘗試了 **Adam** 和 **SGD**，但是不知道為何，理論上明明應該有優勢的 **Adam** 的結果並沒有如預期的凸出，所以最終結果我還是使用了 **SGD**。

### - Batch size:

有一種說法是，大的 **batch size** 會較易收斂，而小的有益於隨機性。我的其中兩次實驗，是 **batch** 為 **64** 與 **32** 的比較，而實驗結果分別是 **0.46000** 與 **0.76083**，可以發現在這兩的模型的狀況下，小的 **batch size** 有較佳的結果。

### - Learning rate:

這次的實驗，我覺得 **learning rate** 佔了很大的一部份，是左右整體表現的重要參數。而我其中幾個實驗是 **learning rate** 設在 **0.002**、**0.005**、**0.015**，實驗結果約是 **0.67375**、**0.81666**、**0.87458**，若直接躍升到 **0.1**，訓練過程會有明顯的震盪，連收斂都有困難。

### - Norm:

出於個人習慣，個人都是設成 **0.9** 左右，而在某次朋友的建議下略改成 **0.92**，疑似有稍微的成長，所以我在最後採用 **0.92**。

- Drop out:

如同大家常使用的，drop out 都是設在 0.1~0.5 左右，我全部的實驗都是設成 0.5，沒有特別作實驗。

b. 最後選擇此值 / 此 scheduling 的原因

最後的值，由上面的順序排下來是(7122, True, SGD, 32, 32, 0.015, 0.92, 0.5)，public accuracy = 0.87458，採用的原因只是因為我的實驗是一種 greedy 的方法，隨著實驗的順序最後採用的結果。當然有落在 local maximum 的問題，我也無從解決。