

1. 請說明你實作的 CNN 模型(best model), 其模型架構、訓練參數量和準確率為何？(1%)

我實做的CNN模型是將原本助教模型的層數再加深一層，並在後面兩個 max pooling 之前再加了兩層，形成 2, 2, 2, 4, 4 個 convolution 層，中間穿插 max pooling，到最後接 fully connected 時還是 [512, 4, 4]。Fully connected 是 [512*4*4] -> [1024*4] -> [1024] -> [512] -> [11]，並在 Linear 之後都有加一個 Dropout(0.5) 來去掉部份資訊，以上就是 model 部份。然後在 optimize 改成 SGD, momentum 設 0.9, learning rate 設 0.05。

以上的架構與參數就能達到在 kaggle 上就能達到 0.855 的分數，在 training 時的 validation accuracy 也能達到 0.81。

2. 請實作與第一題接近的參數量，但 CNN 深度（CNN 層數）減半的模型，並說明其模型架構、訓練參數量和準確率為何？(1%)

因為原本 model 用的架構是 2, 2, 2, 4, 4 層的 Convolution 組合而成的，所以這邊就把層數改成 1, 1, 1, 2, 2，其餘參數像是 fully connected, learning rate, epoch 等都是同樣的。這樣訓練時間雖然也幾乎減半了，但是辨識成功率只有到 0.82785，就算訓練再久也沒辦法再上去了。

3. 請實作與第一題接近的參數量，簡單的 DNN 模型，同時也說明其模型架構、訓練參數和準確率為何？(1%)

這題的實做將前面的 Convolution 都拿掉了，只剩後面的 fully connected，並把圖片的 3*128*238 接到 512*4*4 之後都跟我原本的一樣。這樣的訓練時間比起原本少很多，但是正確率只剩一半 0.48475，明顯降低了很多，沒有 convolution 的提取特徵就使 model 找不到好的參數了。

4. 請說明由 1 ~ 3 題的實驗中你觀察到了什麼？(1%)

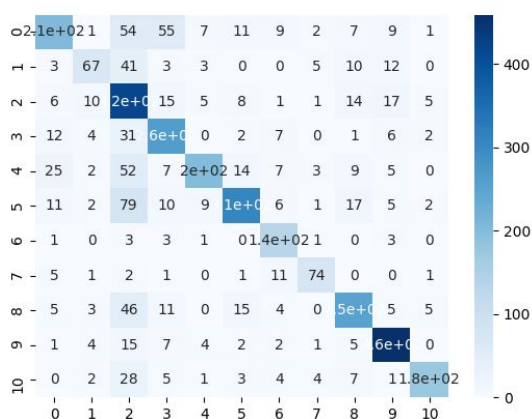
在一開始修改助教原本的模型的時候，發現加深 Convolution 層對於分辨正確率有很好的幫助，但是一想要把 [512, 4, 4] 加深成 [1024, 2, 2] 都沒有得到太好

的結果。所以加深 CNN 層數確實有幫助，雖然 Training 時間變化的很明顯，但是也不能盲目的加深，否則最後留下來的資訊太少反而對於正確率沒有很大的幫助。如果只使用 DNN，沒辦法萃取重要資訊，正確率大大降低，說明前面的 convolution 對於 machine learning 來說非常關鍵，並沒有辦法只是把原本的圖片資訊透過 fully connect 來找到一堆參數就能有好的訓練效果。

5. 請嘗試 data normalization 及 data augmentation，說明實作方法並且說明實行前後對準確率有什麼樣的影響？(1%)

使用 data normalization 時的參數參考了 hw6 用來 normalization 的參數 $\text{mean} = [0.485, 0.456, 0.406]$, $\text{std} = [0.229, 0.224, 0.225]$ 來正規劃，並不是真正的去算實際資料的平均與標準差，這些參數對圖片來說是比較好的。data augmentation 就如同範例 code 中作的一樣，有翻轉及旋轉，並沒有做額外的更改。以上這兩種 preprocessing 的方法有助於減少 overfitting 的狀況，雖然在 training 時的正確率沒有明顯的改變，但是在 kaggle 上的 testing 分數能有一點改善。

6. 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析] (1%)





以上這個 confusion matrix 是用 validation set 做出來的資料，在 label 2 的地方特別容易搞錯，可能是因為他的 dataset 裡是蛋糕或點心，所以可能跟咖啡色的其他食物或是因為白色的盤子而認成其他食物。Label 6 是義大利麵，所以在辨認的時候麵條的形狀對 model 來說非常明顯，所以正確率相對的高出很多。