1. **(2%)** 請說明你實作的 CNN model，其模型架構、訓練參數和準確率為何？並請用與上述 CNN 接近的參數量，實做簡單的 DNN model，同時也說明其模型架構、訓練參數和準確率為何？並說明你觀察到了什麼？

(Collaborators:) 答：

CNN model :

model1	model2	model3
Conv(64,3,3)+BN+relu	Conv(64,3,3)+BN+relu	Conv(64,3,3)+BN+LeakyReLU
Conv(64,3,3)+BN+relu	Conv(64,3,3)+BN+relu	MaxPooling(2, 2)+Dropout(0.1)
MaxPooling(2, 2)	MaxPooling(2, 2)	Conv(128,3,3)+BN+LeakyReLU
Conv(128,3,3)+BN+relu	Conv(128,3,3)+BN+relu	MaxPooling(2, 2)+Dropout(0.1)
Conv(128,3,3)+BN+relu	Conv(128,3,3)+BN+relu	Conv(256,3,3)+BN+LeakyReLU
MaxPooling(2, 2)	MaxPooling(2, 2)	MaxPooling(2, 2)+Dropout(0.1)
Conv(256,3,3)+BN+relu	Conv(256,3,3)+BN+relu	Conv(512,3,3)+BN+LeakyReLU
Conv(256,3,3)+BN+relu	Conv(256,3,3)+BN+relu	MaxPooling(2, 2)+Dropout(0.1)
MaxPooling(2, 2)	Conv(256,3,3)+BN+relu	Conv(512,3,3)+BN+LeakyReLU
Conv(512,3,3)+BN+relu	MaxPooling(2, 2)	MaxPooling(2, 2)+Dropout(0.1)
Conv(512,3,3)+BN+relu	Conv(512,3,3)+BN+relu	
MaxPooling(2, 2)	Conv(512,3,3)+BN+relu	
	Conv(512,3,3)+BN+relu	
	MaxPooling(2, 2)	
	Conv(512,3,3)+BN+relu	
	Conv(512,3,3)+BN+relu	
	Conv(512,3,3)+BN+relu	
	MaxPooling(2, 2)	
Flatten()		
Dense(512)+BN+relu+dropout(0.5)		Dense(512)+BN+LeakyReLU +dropout(0.5)
Dense(512)+BN+relu+dropout(0.5)		Dense(512)+BN+LeakyReLU +dropout(0.5)
Dense(512)+BN+relu+dropout(0.5)		Dense(512)+BN+LeakyReLU +dropout(0.5)
Dense(7)+softmax		
about 7600000 params	about 15500000 params	about 4700000 params
0.70799/0.69824	0.68598/0.68013	0.68124/0.67706

(public score/private score)

Training：

data normalization(全部除以 255)

data augmentation(使用 ImageDataGenerator)

batch_size=128, epochs=250~500

Result：

最後 kaggle 我是以 model1,2,3 各 train 出 3 個 model，最後將 9 個 model
ensemble，得到了 public score: 0.73279 / private score: 0.73307

DNN model：

| Dense(2048)+BN+relu+dropout(0.5) |
|---|
| Dense(2048)+BN+relu+dropout(0.5) |
| Dense(2048)+BN+relu+dropout(0.5) |
| Dense(1024)+BN+relu+dropout(0.5) |
| Dense(512)+BN+relu+dropout(0.5) |
| Dense(7)+softmax |
| about 15500000 params |
| public score: 0.39537/private score: 0.40261 |

Training： 與 CNN 一樣

Result：

與 model2 相比，在接近同樣的參數量下以及同樣的 epochs，DNN 的成效十
分的差。或許將 epochs 拉到很大或是再將參數量加多，DNN 也會有不錯的
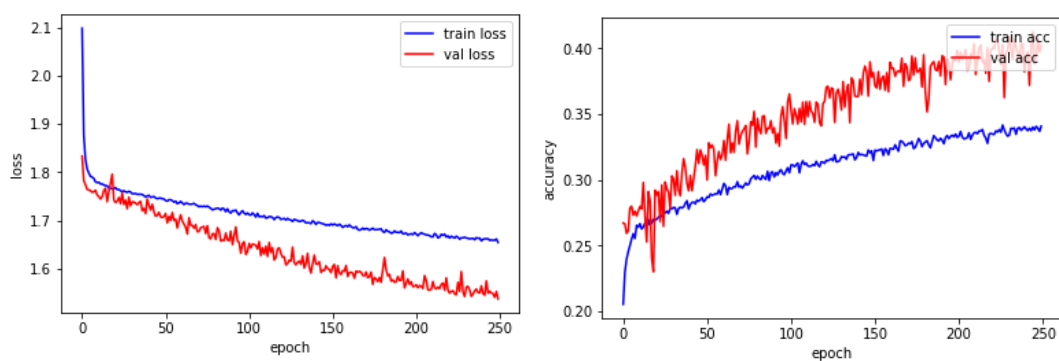成果。不過其所付出的成本就會比 CNN 多很多。

以下討論皆以 model2 為代表

2. (1%) 承上題，請分別畫出這兩個 model 的訓練過程 (i.e., loss/accuracy v.s. epoch)
(Collaborators:) 答 ：

CNN model:
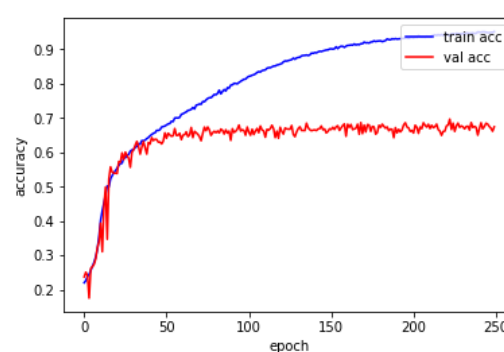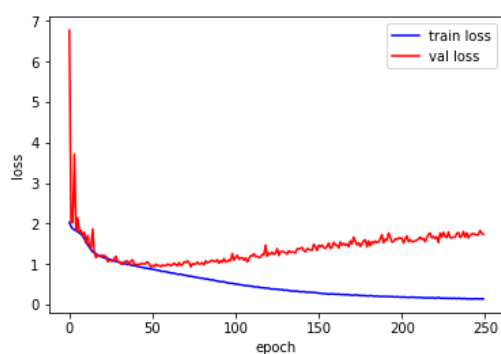
DNN model:

3. (1%) 請嘗試 data normalization, 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？

(Collaborators:) 答 ：

	Public socre	Private score
without data normalization without data augmentation	0.64084	0.62134
with data normalization without data augmentation	0.64112	0.62162
with data normalization with data augmentation	0.68598	0.68013

若未做 data augmentation，雖然 training 時的準確率可達到 99 點多，但從實際結過卻下降很多，可推論應該是發生了 overfitting。

若未做 data normalization，對於整體的準確率並沒有太大的影響。

4. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]

(Collaborators:) 答 ：
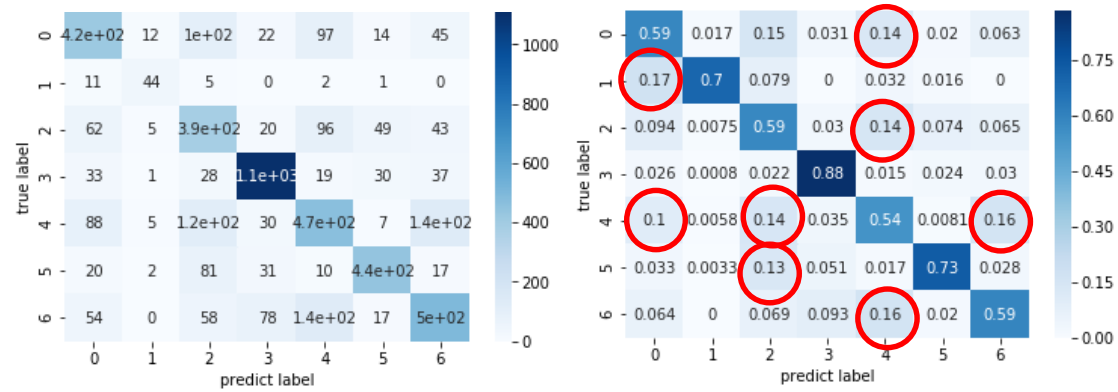
以紅圈圈出大於 0.1 的，可觀察出在 class4(難過)與 class0(生氣)，class4(難過)與 class2(恐懼)，class4(難過)與 class6(中立)，接比較容易用混。