學號：b07901014 系級：電機二　姓名：陳希格

1. 請從 Network Pruning/Quantization/Knowledge Distillation/Low Rank Approximation 選擇兩個方法(並詳述)，將同一個大 model 壓縮至同等數量級，並討論其 accuracy 的變化。 (2%)

   都是經過Knowledge Distillation的相同model，再用經過下列方法
   數量級:都壓成300KB

   Quantization:
   用助教提供的方式(直接改變weight的精度)，但可以改成任意長度。
   壓成9bit 時，大小等於300KB左右，valid accuracy = 0.80。
   大於7bit時 valid accuracy 都在0.8左右，但有趣的是，壓成6bit 時，valid accuracy 掉到了0.21，影響很大。

   Network Pruning:
   用助教提供的方式，Prune rate = 0.9 , epoch = 12 , Prune rate total = 0.28
   Prune完後會重新fine-tune 5 epochs。
   最終valid accuracy = 0.54
   可以看到pruning的accuracy掉了很多，推測其原因重新fine-tune 的epochs不夠多(但其實要很多的話就等於重新Knowledge Distillation了)，或是model已經過於簡單，無法再壓縮。

以下三題只需要選擇兩者即可，分數取最高的兩個。

2. [Knowledge Distillation] 請嘗試比較以下 validation accuracy (兩個 Teacher Net 由助教提供)以及 student 的總參數量以及架構，並嘗試解釋為甚麼有這樣的結果。你的 Student Net 的參數量必須要小於 Teacher Net 的參數量。(2%)
   x. Teacher net architecture and # of parameters: torchvision’s ResNet18, with 11,182,155 parameters.
   y. Student net architecture and # of parameters:
   a. Teacher net (ResNet18) from scratch: 80.09%
   b. Teacher net (ResNet18) ImageNet pretrained & fine-tune: 88.41%
   c. Your student net from scratch: 62.80%
   d. Your student net KD from (a.) 73.67%
   e. Your student net KD from (b.):: 71.69%

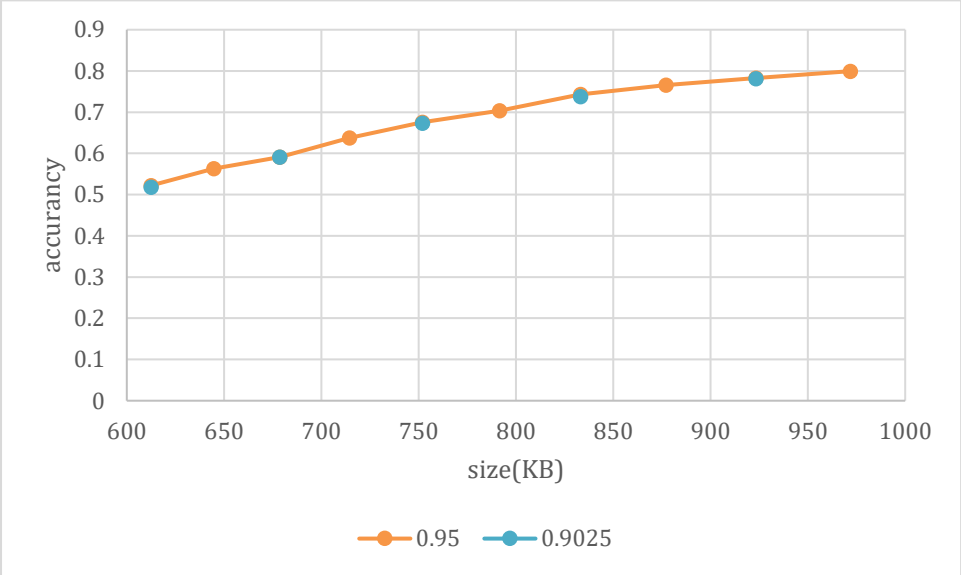
   studentnet架構:原本的(未更動)
   studentnet參數大小: 32K 個

epoch = 40
optimizer =AdamW(lr = 0.01)

可以看到 c明顯較低，符合我們的預測，因為沒有大model的資料。然而d比e還高就比較特別了，代表說大model 的fine-tune 並不一定對小model有幫助。

3. [Network Pruning] 請使用兩種以上的 pruning rate 畫出 X 軸為參數量，Y 軸為 validation accuracy 的折線圖。你的圖上應該會有兩條以上的折線。(2%)

pruning rate = 0.95 pruning rate2 = 0.9025



可以看到每次prune rate 大小其實不太影響結果，兩條線幾乎重和。這可能是因為Prune完後會只重新fine-tune 2 epochs，改變的不多，weight的排序幾乎沒變。

4. [Low Rank Approx / Model Architecture] 請嘗試比較以下 validation accuracy，並且模型大小須接近 1 MB。 (2%)

a. 原始 CNN model (用一般的 Convolution Layer) 的 accuracy
b. 將 CNN model 的 Convolution Layer 換成參數量接近的 Depthwise & Pointwise 後的 accuracy
c. 將 CNN model 的 Convolution Layer 換成參數量接近的 Group Convolution Layer (Group 數量自訂，但不要設為 1 或 in_filters)