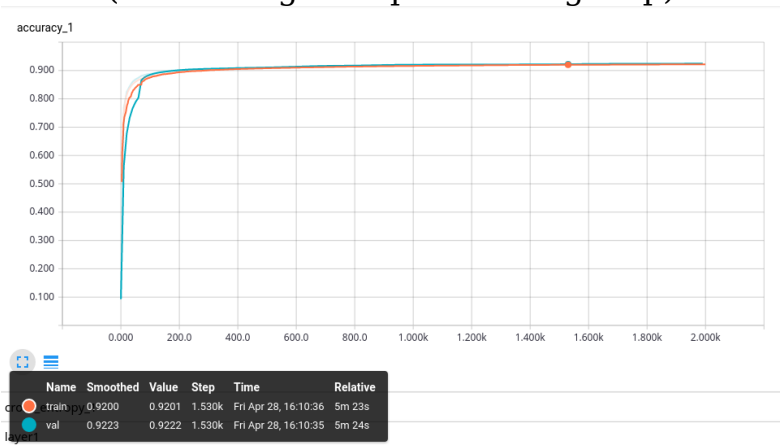
Tensorflow Assignment Report

系級：電機四
姓名：陳品融

Environment: Linux 4.9.11-1-ARCH
Language: Python 3.6.0
Tensorflow version: 1.0.0

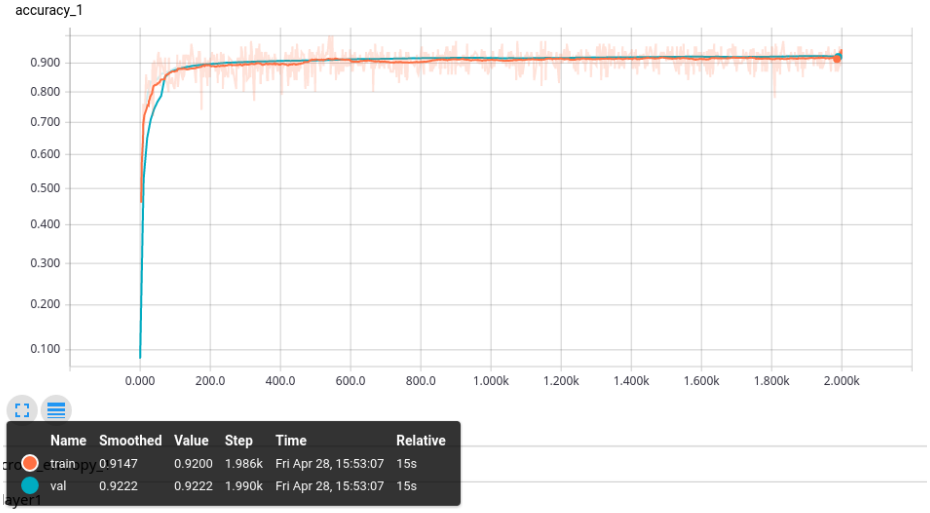
Practice 1:

1. gradient descent (all training data per training step):



   training time: 323 sec
   test accuracy: 0.9216

2. stochastic gradient descent (mini-batch):

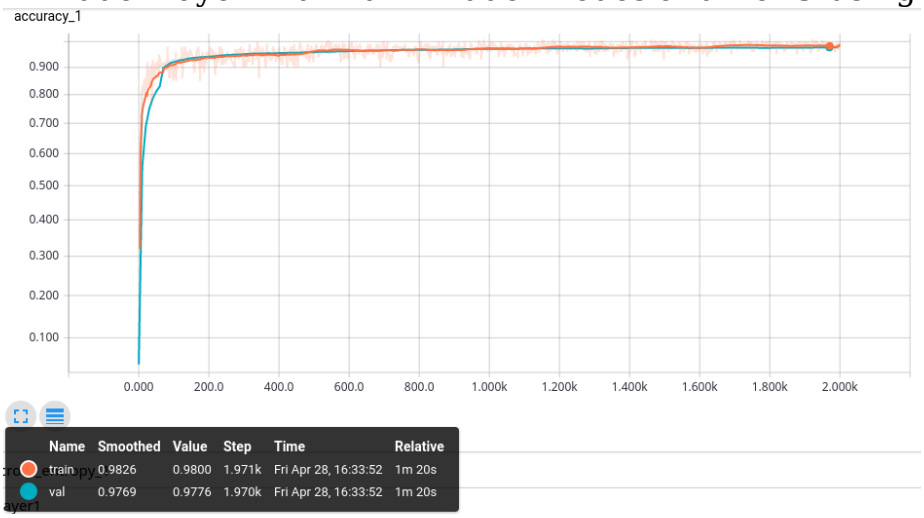


   training time: 15 sec
   test accuracy: 0.9216

3. Observation:
   (1) 此 model 最後 validation accuracy 大概會收斂到 0.922。如果以此來衡量 training time 的話，我們可發現 sgd 的效率大概比原本的 gradient descent 方法快上 20 倍。雖然看似 sgd 所需的 step 比較多，然而由於我的 batch_size 爲 100，所以 sgd 實際上所需的 training data 也是來得較少。
   (2) 以上的圖有經過 smooth 的處理，若沒有 smooth 的話可發現，sgd 的曲線相較之下震盪許多。

Practice 2:

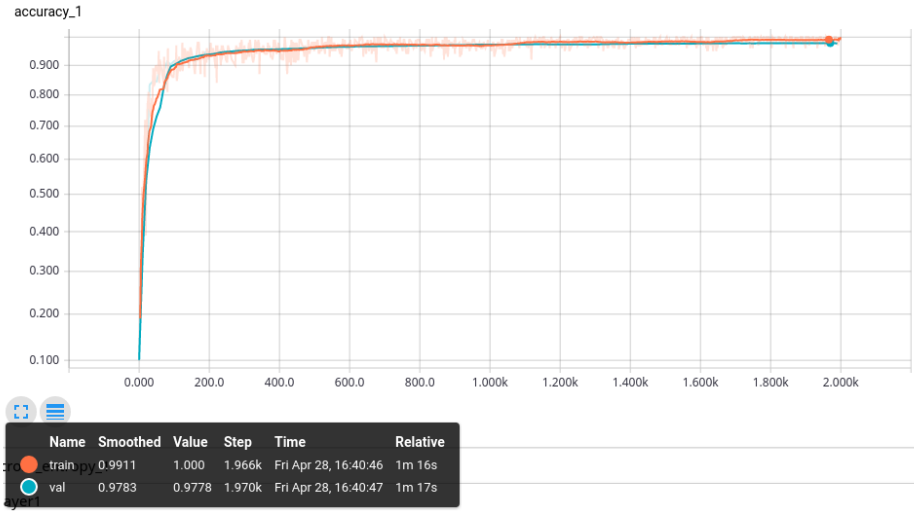
1. model: 1-hidden layer with 1024 hidden nodes and ReLU using SGD



training time: 80 sec
test accuracy: 0.9728

Practice 3:

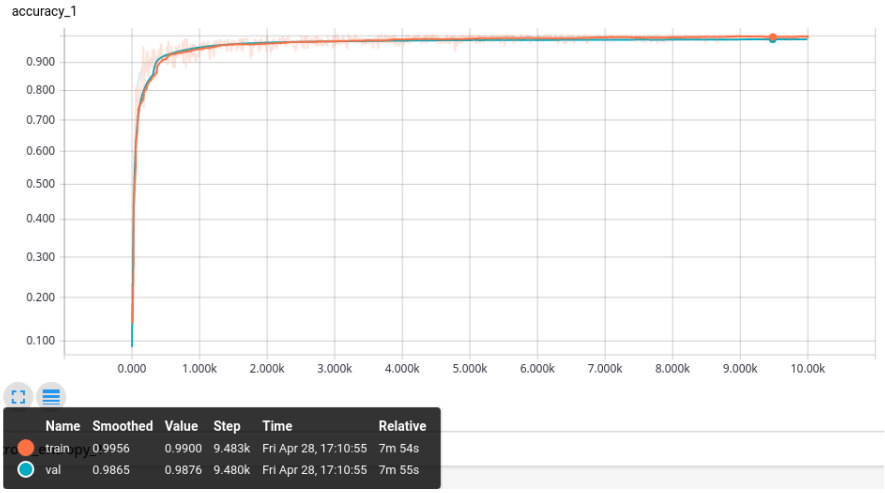
1. model: 2-hidden layer with 1024, 256 hidden nodes respectively, exponential_decay(starter_learning_rate=0.3, decay_steps=100, decay_rate=0.96, staircase=True),
※ decay_learning_rate = starter_learning_rate * decay_rate ^ (global_step / decay_steps)



training time: 76 sec
test accuracy: 0.9745

Practice 4:

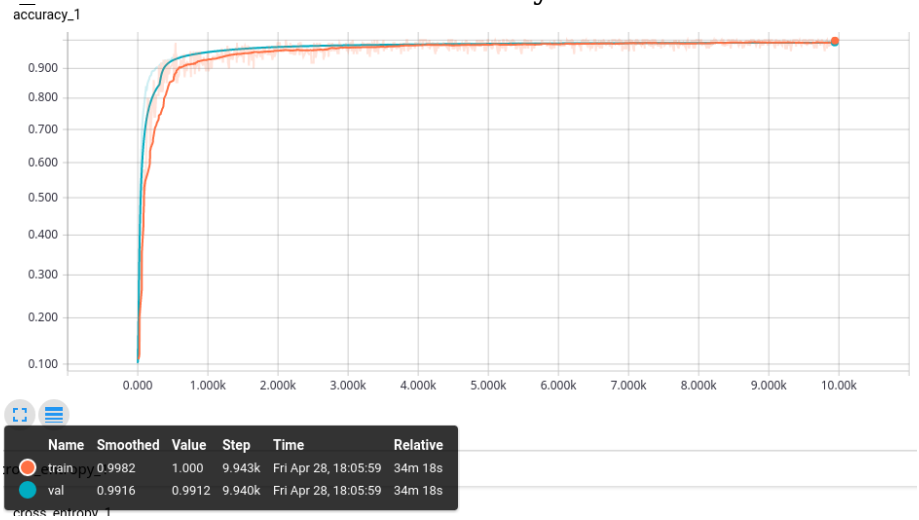
1. model: CNN(Both Conv1 and Conv2: 16@5x5 filters at stride 2, padding='SAME') followed by one fully connected layer using Adam



training time: 475 sec
test accuracy: 0.9875

Practice 5:

1. model: replace the strides in practice 4 by max pooling operation of stride 2, kernel size 2 ,and padding=’SAME’, apply dropout with dropout_rate=0.5 before the readout layer



training time: 2058 sec
test accuracy: 0.9903

Save a checkpoint: 每 100 個 steps 存一次，只留最近的 5 個 checkpoint 檔