DeepLearning HW3

學號:313834004 姓名:周彥宏

Analyze data

這次的Data分成了3個log file. 分別是Training, Test, metrics. 三個log file的內容除了裡面的feature會有不同以外, 格式都是一樣的. 所以對於資料前處理可以使用通用的方法來去實作. 要注意的是每個檔案前4行是有關Setting的資訊, 這部分對於這次畫圖來說, 並不是需要抓取的, 所以會使用判斷是將前面的Setting捨棄掉.

Data preprocessing



首先, 我想要一次讀取三個檔案, 所以把Data全部丟進同一資料夾中, import os module, 並使用os.listdir function來幫我抓取這個資料夾內所有檔案名稱, 變成一個list. 接著我利用一個Dictionary來存三個log file的資料, 值得注意的是每個log file我也設成dictionary來存所有的features.

如同剛剛上面所提到, 我在讀取檔案時, 是一行一行讀取, 然後判斷這行是不是有step來確定並不是前四行的setting line.



接著我對要抓取的資料作前處理, 因為我們不需要epoch的資訊, 所以我選擇先用steps當作切分點把**epoch drop**. 接著切分**’:’** 來存steps的數值, 再**切分左右大括號**把所有feature切下來, 並觀察到中間都是使用**逗號分隔**, 所以又再使用逗號當切分點. 可以注意到我的feature只到[:-1], 是因為三份log file,最後都會多一個逗號, 且是空白. 我直接這樣做, 也不用去多做處理就捨棄他.

接著再讀取feature name時, 我使用到.strip()這個function來幫我去除feature name頭尾的單引號, 因為我是使用冒號當切分點, 把數據跟名稱分開.

Plot



由於我想要一次性畫三張圖, 所以使用subplot幫我在整張畫布上切分成三個圖片. 接著我使用我處理好的資料來畫, 所有的features(除了steps), 都會將steps當成x軸參照後, 畫出線段. 並且給予label(等於是在幫這條線命名), 這樣最後使用plt.legend()時,才會顯示出來. 然後由於範例有給網格線, 所以我也有加上網格線.