

1. (1%) 請說明你是如何 normalize discrete 跟 continuous 的 feature

Ans:

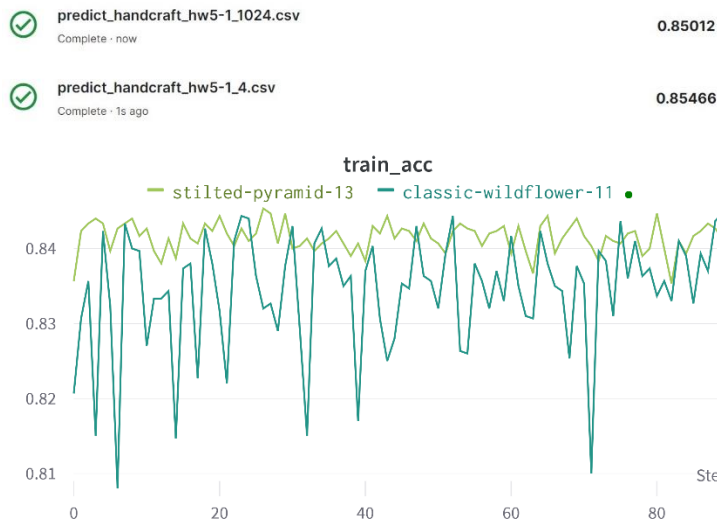
I only used normalization on continuous feature, that is "age", "fnlwgt", "capital_gain", "capital_loss", and "hours_per_week" so that the mean is approximately to zero and standard deviation is approximately to one.

2. (1%) 使用 DNN 做 feature transformation，將 output dimension 設為 4 跟 1024，並丟進 linear SVM 訓練，比較 leaderboard 上的結果，並說明造成這樣結果的原因 (hint: linear SVM 本身是 linear classifier，資料必須是 linearly separable 的資料)

Ans:

The outcome is 0.85012 and 0.85466 for 1024 and 4 output dimension respectively with the same experience setting including normalization. My perspective for the difference of these two result is when you specifically focus on narrow features in this dataset, we can obtain the better outcome.

Still, we can observe the training accuracy record shown as below. The light green line is narrow down to 4 output dimension and it's stable during training loop. On the other hand, 1024 dimension seems too much features to train.



3. (2%) Refer to math problem