

機器學習 Machine Learning LAB 2

系所：電機所

學號：609415074

姓名：蔡承宏

1. 執行說明(Execution description)：

首先我先更改 Dataset.py，我用 mean() 函式把缺失值補上去，之後等到下面 label 跟 train 的資料分好後再用一個叫做 MinMaxScaler() 的函式把 Data 標準化，這部分要特別注意我們函式擺放的位置，因為有資料型態的問題 pd 跟 numpy 的資料型態要先進行轉換，沒有辦法直接通用。

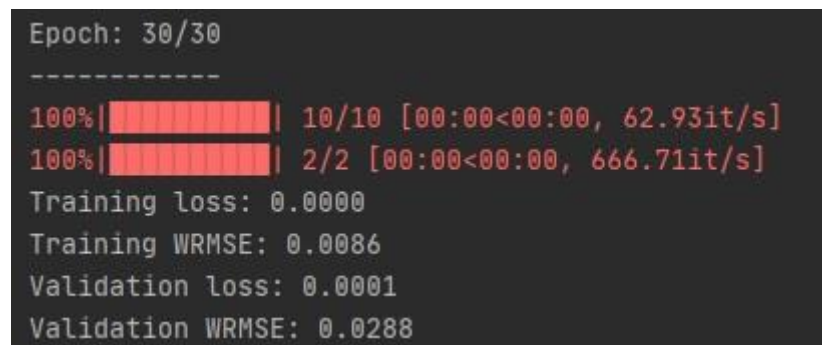
其次我改的是 model.py，我再網路架構裡面進行調整，把她從 1 層改成 3 層，然後要特別注意中間連接層跟最後一層的輸入要改成第一層的輸出，也就是 512，而最後一層的輸出跟接到輸出層的輸入則要設定為 1024，並且要每一層網路架構中間都要加上 Relu 跟 drop out，可以降低 Training 的 Loss 跟 WRMSE，我一開始沒加入時，測試的結果非常差，但是不知道為什麼加入後就變好了，這部分其實我不太了解，希望助教或老師可以上課時可以幫解惑。

接著我做的是 Train.py，這部分不太好解釋，因為改了很多，我先把原本的 dataset 用一個叫做 train_test_split 的函式把 Dataset 切成 9:1，training 的 dataset 是 9，validation 的 dataset 是 1，之後只需要再對 validation 的驗證進行 coding 即可，然後我都是在 for(e) in range(epoch) 的迴圈裡面撰寫，最後再畫圖將結果一並輸出即可。

最後改的是 test.py，更改的過程跟 Dataset.py 一樣，所以我就不贅述了。

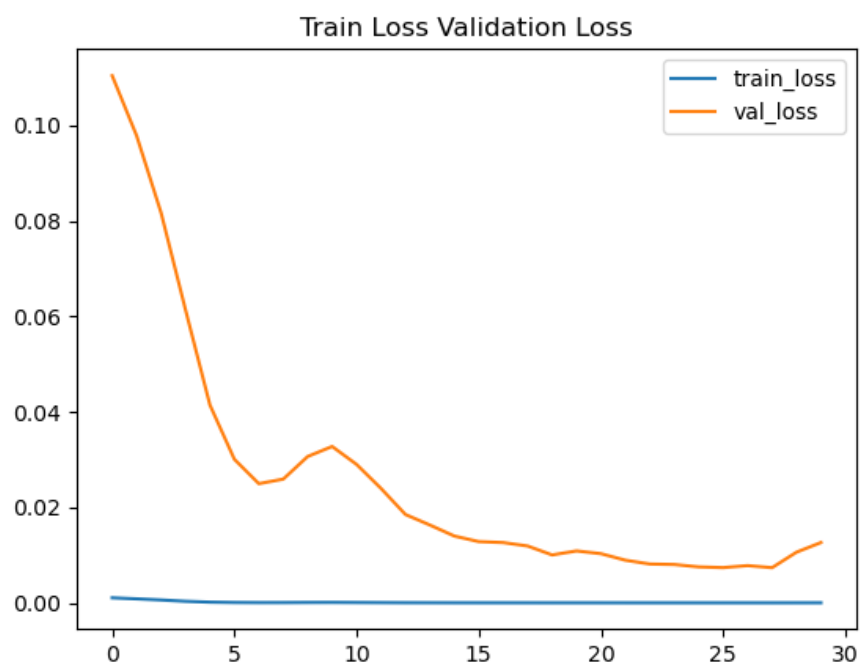
2. 實驗結果(Experimental results)：

這次的實驗結果，如下圖

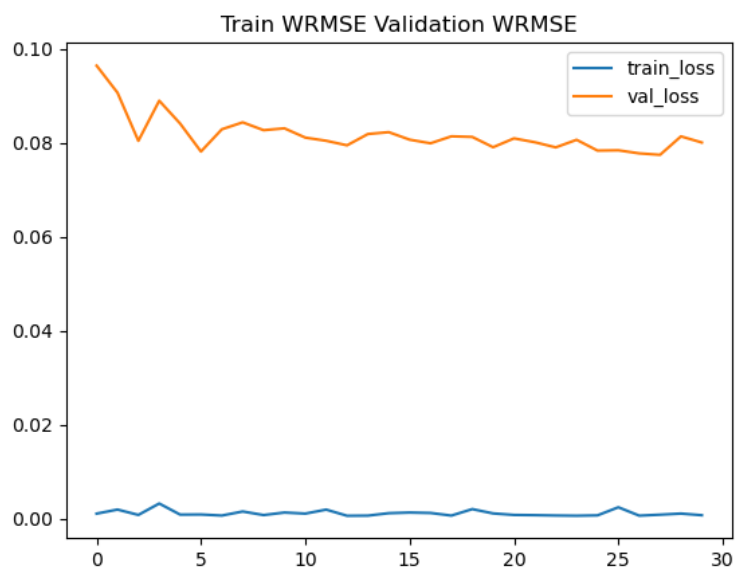


```
Epoch: 30/30
-----
100%|██████████| 10/10 [00:00<00:00, 62.93it/s]
100%|██████████| 2/2 [00:00<00:00, 666.71it/s]
Training loss: 0.0000
Training WRMSE: 0.0086
Validation loss: 0.0001
Validation WRMSE: 0.0288
```

圖一、Training loss & Training WRMSE & Validation loss & Validation WRMSE



圖二、Training 跟 Validation 的 Loss



圖三、Training 跟 Validation 的 WRMSE

3. 結論(Conclusion)：

這次的作業我卡最久的地方是 `Dataset.py`，因為我一開始不知道為什麼資料沒辦法給 `Model` 吃，到後來我才注意到資料型態的部分，我的標準化一直沒有辦法輸入，常常 `Training` 的結果都是 `nan`，到最後才找到問題點。

4. 討論(Discussion)：

這一次的作業讓我學到了很多，主要都是偏資料結構的部分，讓我學到了很多，因為遇到了很多的問題，像是一開始的怎麼增加 **hidden layers**、怎麼切割 **dataset**、怎麼平均值、標準化...等，這些我全都不會，我一步步的慢慢找資料、問人，最後才做出來這次的作業，這次的作業難易度我覺得剛剛好，不會太難也不會太簡單，讓我學到了很多東西。^__^