National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: 盧育琦　　　　　Student ID: 112034582

Due on 2024/05/02.

Note: DO NOT exceed 3 pages.

1. (15 points) Experiment with different window sizes and steps. Train the model using **3** different combinations of window size and step. Evaluate the Mean Squared Error (MSE) for each configuration. Report the MSEs using a table and analyze the results. (Approximately 100 words.)

Window size	Step size	Train loss	Val loss	Best Val loss
10	15	114.0494	252.0068	252.0068
25	10	108.4538	56.2507	56.2507
5	10	35.5805	60.9022	60.9022
50	10	47.9902	19.2300	19.1380
50	5	16.5269	14.3305	14.3305
25	5	7.1311	16.9359	16.9359

Epochs=100

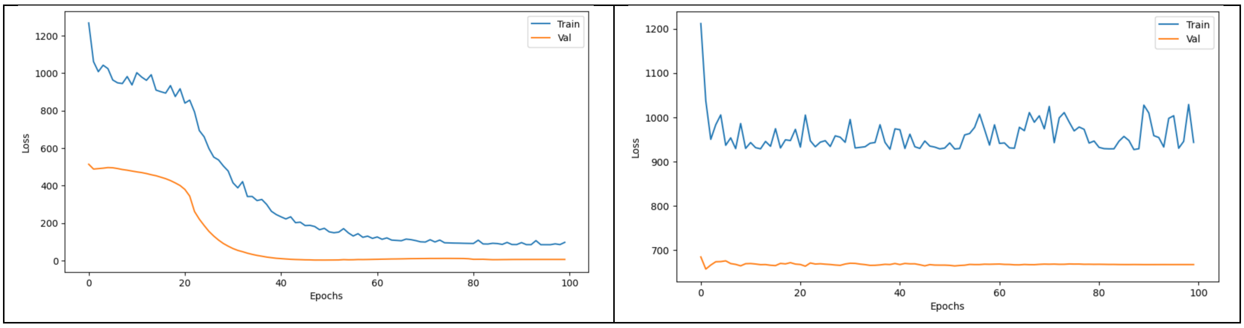
從我的紀錄中可以發現當 step 越小的時候，loss 也會變得更小，但似乎無法從 step size 是否大小於 window size 找出相關性。

2. (Approximately 200 words.)

(i) (15 points) Include 'Volume' as an additional input feature in your model. Discuss the impact of incorporating 'Volume' on the model's performance.

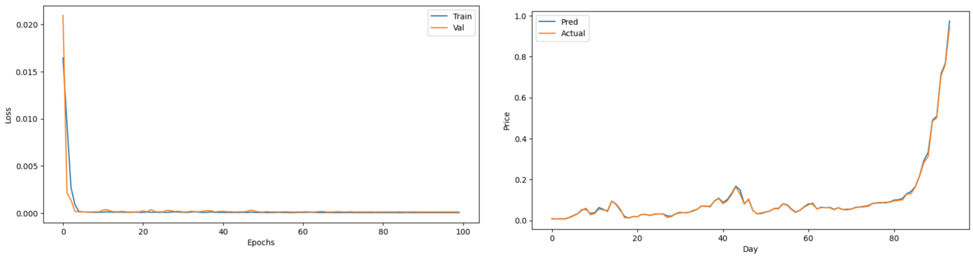
從下面的比較圖我們可以看到包含‘Volume’的 loss 會變很大，那是因為在原始檔案中的 Volume 值都非常的大，從而導致比較大的變異。

Exclude ‘Volume’	Include ‘Volume’
Epoch 100/100,	Epoch 100/100,
Train loss: 97.0559,	Train loss: 943.4127,
Val loss: 5.7606,	Val loss: 667.7104,
Best Val loss: 2.3886	Best Val loss: 657.5481

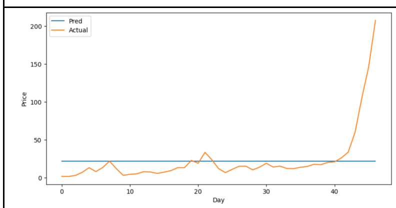
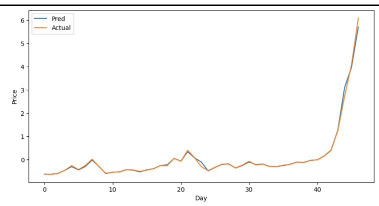
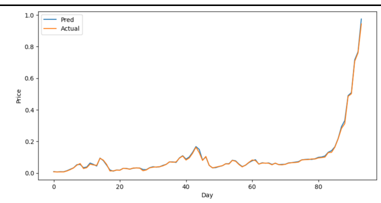


(ii) (15 points) Explore and report on the best combination of input features that yields the best MSE. Briefly describe the reasons of your attempts and analyze the final, optimal input combination.

最佳特徵組合能夠最大程度地降低 MSE loss，在我的嘗試中，Input 所有 Features 的 loss 使模型有最好的表現，我查到藉由更多的數據可以使模型更有效的去預測結果，但有時候也要根據要預測的目標決定特徵的選擇。另外我發現只要再加上 Normalized ，就能使模型有最佳的 MSE。



3. (15 points) Analyze the performance of the model with and without normalized inputs in Lab 4. You can use experimental results or external references (which must be cited) to support your conclusions on whether normalization improves the model's performance. (Approximately 100 words.)

Without Normalized Epoch 100/100, Train loss: 805.2155, Val loss: 1314.7269, Best Val loss: 1307.7554	**With** Normalized (StandardScalar) Epoch 100/100, Train loss: 0.0014 Val loss: 0.0014, Best Val loss: 0.0007	**With** Normalized (MaxAbxScalar) Epoch 100/100, Train loss: 0 Val loss: 0.0001, Best Val loss: 0.0001
		

從上表的比較可以看到，透過 normalized 可以有效地優化表現，我使用了兩種 SKlearn 模型，當中又以 MaxAbxScalar 的表現更好，它將數據歸一化到[-1,1]的範圍，並保留原始數據的符號。

4. (10 points) Why should the window size be less than the step size in Lab 4? Do you think this is correct? If you use external sources, please include references to support your response. (Approximately 50 words.)

   我認為是不需要的，從第一題的回答中，可以看到「window size」與「step」對於結果並沒有直接的關聯，因此可能可以推斷窗口大小和步長在 LSTM 模型中是獨立的參數。

5. (15 points) Describe one method for data augmentation specifically applicable to time-series data. Cite references to support your findings. (Approximately 100 words.)

   - 時間序列差值法：通過在現有數據點之間進行插值來生成新的數據點，從而擴充數據集。插值技術可以根據現有數據的特性，如趨勢和季節性，來生成具有相似特性的新數據點，從而有效地增加數據的多樣性。
     - 參考文獻: Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nded.). OTexts. https://otexts.com/fpp2/regression.html

6. Discuss how to handle window size during inference in different model architectures (approximately 150 words)

   (i) (5 points) Convolution-based models：通常需要使用滑動窗口方法，將 input data 分解成窗口，然後將每個窗口傳遞到模型中進行預測。通過調整窗口的大小和步長，可以控制 input data 的覆蓋範圍，從而影響模型的預測結果。

   (ii) (5 points) Recurrent-based models ：通常需要使用固定窗口大小的方法，將模型設置為接受固定大小的 input sequence，並將每個序列傳遞到模型中進行預測。這樣可以確保模型的預測與訓練期間一致。

   (iii) (5 points) Transformer-based models：通常不需要對窗口大小額外進行處理，它可以處理任意長度的序列，可以直接將整個時間序列作為模型的輸入。