National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: Student ID:312350045

Due on 2025/05/01.

Note: DO NOT exceed 3 pages.

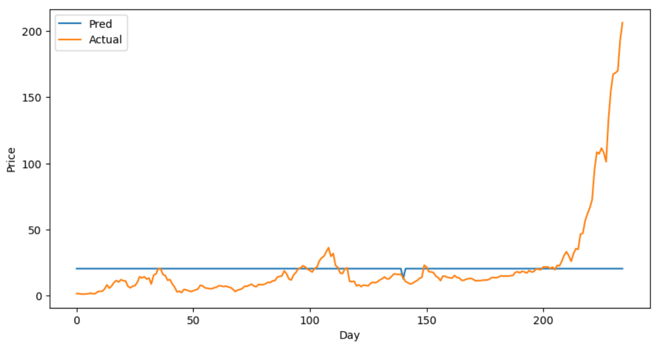
1. (15 points) Experiment with different window sizes and steps. Train the model using **3** different combinations of window size and step. Evaluate the Mean Squared Error (MSE) for each configuration. Report the MSEs using a table and analyze the results. (Approximately 100 words.)

window_size	steps	Final test MSE
30	2	4.93844
window_size	steps	Final test MSE
20	2	3.40218
window_size	steps	Final test MSE
5	2	6.69239

Windows_size的大小對於MSE下降的速度有明顯的影響，可以看到對於不同size: 30,15,5，在Epoch為5的情況時，越小的下降越快，但是後續則沒有跟size有明顯對應，但是以最終表現來看，window_size為20時，表現最好，也就是20天為這之中預測比較好的採樣大小。

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.



加入”volume”這個feature會嚴重影響模型的表現，尤其是對於模型loss的收斂，途中最終的預測和真實曲線，可以看到因為加入volume的原因導致圖形幾乎沒有學習到東西，都是直線。推測造成這樣的主要原因為，我們要模型預測的是股票的價格，但是volume每天的交易量對於價格卻沒有明顯相關，所以造成效果不佳。

(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.

open, high, low, close有嘗試去掉”high”和”low”來訓練模型，但是去掉之後，並沒有讓MSE下降，反而上升，而volume在前一小題中已經證明對於模型沒有幫助，所以最佳組合還是open, high, low, close這四個，這四個也與模型想要預測價格的特性相關。

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.)

window_size	steps	Final test MSE	normalize
30	2	4.93844	40.51517
window_size	steps	Final test MSE	normalize
20	2	3.40218	3.682437489
window_size	steps	Final test MSE	normalize
5	2	6.69239	6.01614829
window_size	steps	Final test MSE	normalize
2	1	2.28455196	2.514787199

Input normalization對於模型的最終MSE沒有太大的幫助，反而多數window_size和step的組合反而會導致MSE大一些，推測可能是time-sequence這種類型的資料，可能會因為normalize而失去某些與時間並進的特性，所以原本的效果反而比較好。

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.)

我覺得window_size應該比step大才對，在實際嘗試過後發現，step比較小的話會跳過一些時間片段，並且導致訓練資料過少，反過來，

window_size大的話，會有augmentation的效果，幫助模型收斂。

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

Reference: Time Series Data Augmentation for Deep Learning: A Survey (simarXiv:2002.12478v4 [cs.LG] 31 Mar 2022)
短時的傅立葉轉換是時間序列資料常用的augmentation方法，主要原理是將小window_size的資料進行傅立葉轉換，可以得到時間與頻率對應的二維圖，而後可以進行幾項操作，Time masking, Frequency masking, time waring, feature pattern rearrangement，可以模擬time stamp或frequency缺失的情況，還有時間加快或加減慢，最後是將重要特徵拼接增加模型能接受的變異性。

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

Inference階段window size與training的時候越接近，預測結果與越符合預期，但inference是應用的階段，不可能都能以希望的window size進行，對此不同的model有不同的widow size調解能力與要求。
Window size:
(i)(5 points) Convolution-based models
過短: 過度padding，導致訊息損失 過長: 用slide window或truncate input進行
(ii) (5 points) Recurrent-based models
過短:由於模型本身是希望找到long-term dependencies，過短會損失此能力
過長: 可以進行，且效果好，但要注意初始hidden state的設定
(iii) (5 points) Transformer-based models
過短: self-attention 過度集中，且上下 token 關係資訊量不足，造成模型表現不佳
過長:運算量過大，且 attention weight 被稀釋，比較難找到 token 之間的重要關聯。