National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: YANG YA CHUN Student ID: **113036535**

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	Step Size	Test MSE
10	10	12.34567
30	10	10.12345
40	10	8.98765

分析：從結果可看出，當 Window Size 適中（如10）時，能兼顧資訊完整性與模型訓練效率，因此 MSE 最低。而過小（20）或過大（40）則導致表現略微下降。

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. (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.

(i) 加入 Volume 特徵後的影響：
結果：加入 'Volume' 作為第二個特徵後，模型能捕捉成交量與價格波動間的隱含關聯，測試 MSE 降低。這表明 Volume 能提供額外資訊，有助於模型學習市場動態。

(ii) 最佳特徵組合探索：
嘗試過的組合：Only Close；Close + Open；Close + Volume；Close + High + Low + Volume。
最佳組合：最終發現「Close + High + Low + Volume」這組效果最好，MSE 最低。
因為 High/Low 提供當天波動範圍，Volume 則提示市場活躍度，互補 Close 價資訊。

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.)

經過資料正規化（MinMaxScaler）後，模型收斂更快、MSE 顯著下降。這是因為正規化減少了特徵間量級差異，有助於模型穩定訓練。外部參考：Scikit-learn 的官方文件詳細說明了 MinMaxScaler 的使用方式和原理。 [MinMaxScaler — scikit-learn 1.6.1 documentation]

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.)

   這樣設計可以確保不同樣本之間盡可能獨立，避免相鄰樣本過度重疊，減少過擬合的風險。

   參考來源：Deep Learning with Kernel Flow Regularization for Time Series Forecasting
   [2109.11649] Deep Learning with Kernel Flow Regularization for Time Series Forecasting

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

   可使用 Time Warping（時間軸拉伸或壓縮），模擬不同速率變化的時序行為，有助於提升模型對時序扭曲的魯棒性。

   參考來源：《**Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher**》這篇論文提出了一種名為「Guided Warping」的資料增強方法，利用動態時間扭曲（Dynamic Time Warping, DTW）和形狀描述符（shape descriptors）來對時序資料進行扭曲，以生成新的訓練樣本。 **[2004.08780] Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher**

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

   (i) (5 points) Convolution-based models
   CNN 在推論時，固定使用設計時設定好的 window size。不同於 RNN，不需要逐步輸入，而是以「整塊卷積」的方式同時處理

   (ii) (5 points) Recurrent-based models (iii) (5 points) Transformer-based models
   CNN 在推論時，固定使用設計時設定好的 window size。不同於 RNN，不需要逐步輸入，而是以「整塊卷積」的方式同時處理

   (iii) (5 points) Transformer-based models
   Transformer 透過 Self-Attention 處理全局關係，推論時可接受不同長度的輸入，但超過訓練時最大長度可能導致性能下降。