National Tsing Hua University

11320IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: 魏詠新 Student ID: 113034542

Due on 2025/05/01.

Note: DO NOT exceed 3 pages.

(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 | Train loss (MSE) | Val loss(MSE) |
|-------------|---------|------------------|---------------|
| 10(原本的) | 15(原本的) | 96.7729 | 56.9863 |
| 30 | 10 | 63.6139 | 48.7069 |
| 20 | 5 | 7.4623 | 8.4877 |
| 15 | 5 | 4.6600 | 11.6603 |

在window=10,step=15可能因模型看到的資訊太少、更新又不頻繁,學不到有效模式,而window=30,step=10,window 太大反而讓模型太難學,因此loss都非常高。在window=15,step=5表現不錯,但在window=20,step=5時表現最佳,能兼顧訓練準確與泛化能力。可見調整 window size 與 step 會大幅影響模型表現,但過小的 window size 或太大的 step,會讓模型學不到足夠的時間序列訊息。

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.

| 差別 | Window/size | Train loss(MSE) | Val loss(MSE) |
|------------|-------------|-----------------|---------------|
| 原本的 | 20/5 | 0.0016 | 0.0017 |
| 加入'Volume' | 20/5 | 0.0014 | 0.0026 |

因為 'Volume' 欄位的資料與其他尺度相差較大,發現須先將特徵及label進行標準化 (使用 StandardScaler),為了與原本的相比,也將原先 Open, High, Low, Close 組合進行標準化,故 loss值不同。由結果可以發現加入'Volume'後在Train loss表現比原本較佳,但 Val loss表現略差一些。

(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.

| 差別 | Window/size | Train loss(MSE) | Val loss(MSE) |
|---------------------------|-------------|-----------------|---------------|
| High \ low \ close \ open | 20/5 | 0.0016 | 0.0014 |
| High \ low \ close \ | 20/5 | 0.0014 | 0.0026 |
| open · Volume | | | |
| High \ low \ close \ | 20/5 | 0.0011 | 0.0033 |
| open · OpenInt | | | |
| High \ low \ close \ | 20/5 | 0.012 | 0.0040 |
| open · Volume · OpenInt | | | |

除了前面做的兩種,增加考慮 OpenInt,因為它可能可以反應市場氣氛。但發現最佳的 MSE(以 val loss 評估)出現在僅使用價格資訊: High, Low, Close, Open 的情況, Val Loss 為 0.0014。雖然加入 Volume 或 OpenInt 會讓Train loss 略微下降,但驗證 loss 卻明顯上升,代表模型在這些情況下出現了 overfitting。這可能代表特徵數量增加,不一定會帶來更好的效果,反而可能因資訊冗餘,導致模型泛化能力下降。

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 | Step | Val loss |
|-------------|------|----------|
| 20 | 5 | 8.4877 |
| 20 | 5 | 1.4828 |

*此部分為比較是否應進行正規化,將正規化後的 Val loss 結果反轉呈現如表格根據實驗結果,模型在未進行正規化時的驗證損失(Val loss)為 8.4877,而經過MinMax 正規化後下降至 1.4828。應是LSTM 模型對特徵尺度較敏感,導致未正規化時模型較難有效學習。可見正規化有助於提升模型收斂速度並穩定訓練過程,讓各特徵對學習的貢獻更平均,因此大幅提升效能。這也與文獻The Investigation of LSTM-Random Search with Various Standardization and Normalization Technologies中對於價格穩定的股票,使用 MinMaxScaler 能顯著提高預測準確度的結論相符。https://www.researchgate.net/publication/379530575 The Investigation of LSTM-Random Search with Various Standardization and Normalization Technologies

- 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.)
 根據在第1題的測試結果發現,在設定window size大於step時表現會較佳。推測應是在股票價格預測中,使用視窗大小大於步長的滑動視窗方法,重疊的視窗提供了更多的上下文信息,有助於模型更好地捕捉時間序列中的模式。
- 5. (15 points) Describe one method for data augmentation specifically applicable to timeseries data. Cite references to support your findings. (Approximately 100 words.)

抖動(Jittering)是時間序列資料增強一種常見的方法。此技術透過在原始數據中添加 隨機噪聲,模擬現實世界中的微小變化,提升模型的泛化能力。具體而言,抖動通常涉 及在每個時間點添加平均值為零的高斯噪聲。這樣的處理可以增加訓練數據的多樣性, 幫助模型更好地適應實際應用中的數據變化,並降低過擬合的風險。

https://maddevs.io/writeups/basic-data-augmentation-method-applied-to-time-series/

- 6. Discuss how to handle window size during inference in different model architectures (approximately 150 words):
 - (i) (5 points) Convolution-based models 推論時可以使用與訓練時相同的窗口大小,以確保特徵提取的一致性。如果需要處理不同長度的序列,可以採用填充(padding)策略維持輸入維度,或使用全局池化層來處理可變長度輸入。
 - (ii) (5 points) Recurrent-based models LSTM等循環模型在推論時對窗口大小較為敏感。這些模型依賴於隱藏狀態來捕捉時間序列的長期依賴關係,因此推論時必須保持與訓練相同的窗口大小,通常採用滑動窗口策略。
 - (iii) (5 points) Transformer-based models 由於自注意力機制允許模型直接學習序列中不同位置間的關係,推論時可以靈活處理不同長度的序列。然而,標準Transformer的計算複雜度與序列長度呈平方關係,因此長序列推論時通常需要分段處理或使用稀疏注意力技術。

https://victorleungtw.com/zh/2025/03/22/attention/