National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name:單允辰　　　　　　　Student ID:112034518

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.)

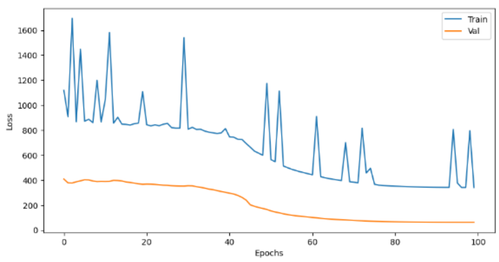
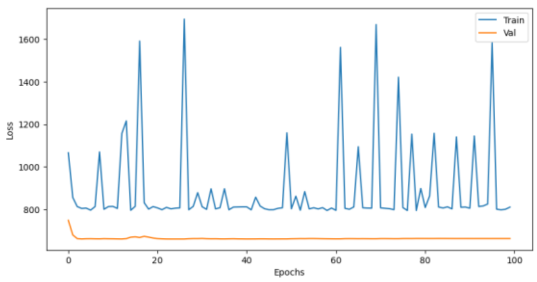
Epochs = 100

Window_size	step	Train MSE loss	Val MSE loss	Best Val loss
10	15	88.8260	268.2098	268.2098
10	30	662.6945	671.0568	671.0568
20	15	189.0113	783.5053	783.5053
20	30	305.1436	466.2636	466.2636

當window size上升時，Train MSE loss下降，不過Val MSE loss反而出現上升的趨勢，可能是模型發生overfitting的狀況。在以上四種組合中，Window_size = 10, step =15模型的Train MSE los 與 Val MSE loss皆為最低，代表此模型表現最好。

2. (Approximately 200 words.)

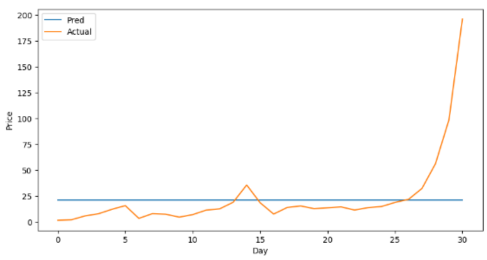
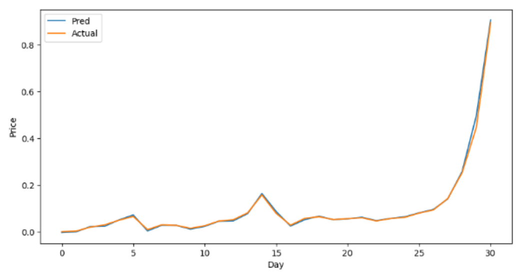
(i) (15 points) Include 'Volume' as an additional input feature in your model. Discuss the impact of incorporating 'Volume' on the model's performance.

features = df[['Open', 'High', 'Low', 'Close']]	features = df[['Open', 'High', 'Low', 'Close','Volume']]
Epochs: 100	Epochs: 100
Train loss: 341.1986	Train loss: 1156.9237
Val loss: 62.8093	Val loss: 602.6532
Best val loss: 62.8093	Best Val loss: 554.1595
	

從上表可以發現，加入Volume後的模型Train loss, Val loss, Best Val loss皆會變高，原因來自於Volume 的數值較高導致，因此若將Volume normalize會使模型表現更好。

(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.
最佳特徵組合可以使模型的MSE較低，在所有模型中，選擇'Open', 'High', 'Low', 'Close','Volume' feature的模型表現最好，因為越多feature可以提供模型更多資訊，另外對資料normalize 會使模型有更低的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.)
從上表可以發現，normalize後的模型表現較好，我使用的是scikit-learn中的MinMaxScaler，將資料縮放到指定範圍，通常是 [0, 1] 區間。

<table>
<tr>
<td>Without normalize
Epoch 100/100
Train loss: 779.3020
Val loss: 756.4386
Best Val loss: 756.4331

</td>
<td>Without normalize(MinMaxScaler)
Epoch 100/100
Train loss: 0.0000,
Val loss: 0.0001,
Best Val loss: 0.0001

</td>
</tr>
</table>

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.)
我認為不正確，因為在LSTM中window size與step size看起來相互獨立，因此沒有一定要限制window size要小於step size。

5. (15 points) Describe one method for data augmentation specifically applicable to time-series data. Cite references to support your findings. (Approximately 100 words.)
對於時間序列資料的資料增強，一種方法是視窗切片(window slicing)，它透過從單

一時間序列中提取固定長度的重疊或非重疊視窗來建立多個樣本。 此技術可以增加訓練資料的多樣性並提高模型的泛化能力（Cui et al., 2020）。

Reference:

Multi-Scale Convolutional Neural Networks for Time Series Classification

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

(i) (5 points) Convolution-based models

在Convolution-based models中,窗口大小對模型預測是有影響的。一般來說,在推理時使用的窗口大小應與訓練時保持一致。但如果輸入序列長度與訓練時不同,則需要相應地調整窗口大小。一種方法是使用可分離卷積,允許不同長度的輸入。另一種選擇是在推理時使用重疊窗口和步幅,從而產生多個預測,然後將它們組合起來。

(ii) (5 points) Recurrent-based models

對於Recurrent-based models,通常可以一次性處理整個輸入序列,而無需考慮窗口大小。但是,對於較長序列,可能需要使用序列截斷或反向傳播通過時間(BPTT)等技術來減輕計算負擔。

(iii) (5 points) Transformer-based models

對於基於Transformer的模型,通常使用固定長度的輸入序列,且在推理時也應保持一致。如果輸入序列過長,可以使用滑動窗口或截斷等策略將其分割成多個片段,分別進行推理並合併結果。另外,一些變體模型(如Reformer)可以有效處理較長序列。