National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

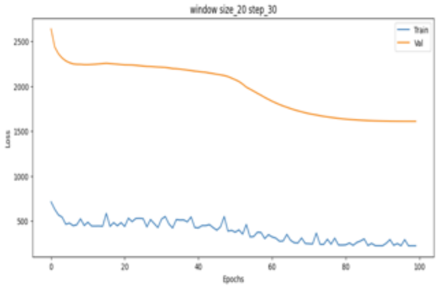
Homework 4

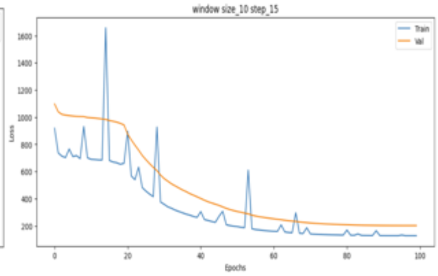
Name: 黃家興　　　　　　　　Student ID: 111003806

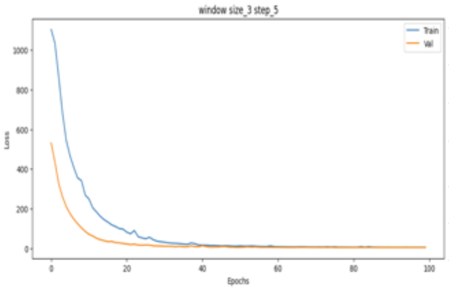
Due on 2024/05/02.
Note: DO NOT exceed 3 pages.

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	MSE
Exp. 1	20	30	126.3445
Exp. 2	10	15	12.3024
Exp. 3	3	5	0.7085



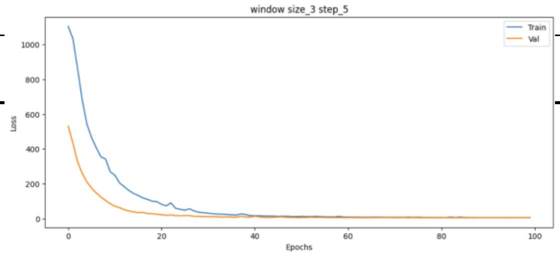


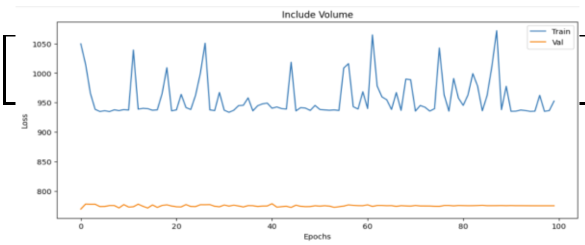


The experiment result show the MSE lose can get better when the window size and step size smaller. Smaller window sizes mean fewer parameters in the LSTM network. That can help capture these local dependencies more effectively and prevent overfitting.

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.

	Feature	MSE		Feature	MSE
Exp. 1	Open, High, Low, Close	0.7085	Exp. 2	Open, High, Low, Close, Volume	38.2546





The performance of model gets worse after add 'Volume' that MSE change from 0.7085 to 38.2546. That should be 'Volume' may not provide meaningful information for the this task or might have a different scale and distribution compared with other features.

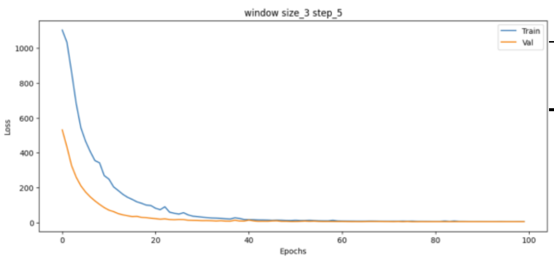
(ii) (15 points) Explore and report on the best combination of input features that yields the best MSE. Briefly describe the reasons of your attempts and analyze the final, optimal input combination.

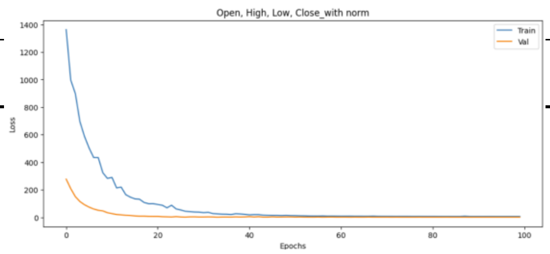
	Feature	MSE
Exp. 1	Open, High, Low, Close, Volume	38.2546
Exp. 2	Open, High, Low, Close	0.7085
Exp. 3	Open, Close, Volume	37.976
Exp. 4	Open, Close	0.7535

From the experiment result show that can get better MSE if exclude ’Volume’. The main features influencing stock price prediction are "Open, Close, High, Low", while 'Volume' is not the main feature, and the most important ones are 'Open' and 'close'.

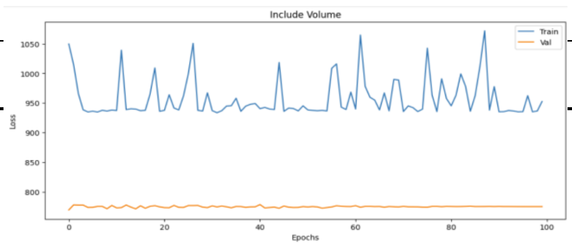
3. (15 points) Analyze the performance of the model with and without normalized inputs in Lab 4. You can use experimental results or external references (which must be cited) to support your conclusions on whether normalization improves the model's performance. (Approximately 100 words.)

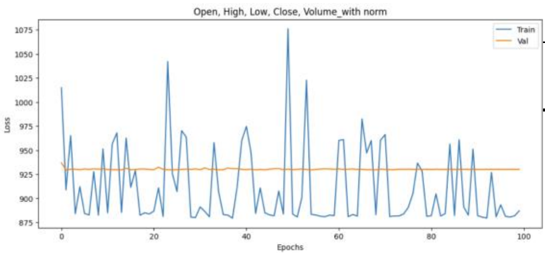
	Feature(without normalized)	MSE		Feature(with normalized)	MSE
Exp. 1	Open, High, Low, Close	0.7085	Exp. 2	Open, High, Low, Close	1.5259





	Feature(without normalized)	MSE		Feature(with normalized)	MSE
Exp. 3	Open, High, Low, Close, Volume	38.2546	Exp. 4	Open, High, Low, Close, Volume	37.9604





Above result shows the performance of the model does not show significant improvement, whether with normalized inputs or without normalization, there could be the model architecture might not be sufficiently complex to capture the underlying patterns in the data. Increasing the model complexity, such as adding more layers or units, could help improve performance. The other possible reason is the hyper parameter might not be appropriately tuned. That need to do the further investigate to identify the factors of this issues.

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.)

   In LSTM models, the relationship between window size and step size depends on the specific requirements of the task and the characteristics of the data. There's no strict rule that mandates the window size to be less than the 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.)

   One method for data augmentation in time-series data is (Dynamic Time Warping) DTW Barycenter Averaging. That is an iterative algorithm that applies dynamic time warping but aligns all the series with an evolving average. Applying this concept, signal averaging will not simply depend on the average of the input signal, but by the magnitudes of each point, the 'weight', in the signal.[1]

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

   (i) (5 points) Convolution-based models

   During inference in Convolution-based models for time series prediction, a fixed-size window is slid across the input time series. At each position, the model applies convolutional filters to extract features and make predictions. This sliding window approach enables capturing temporal patterns and forecasting future values based on historical data.

   (ii) (5 points) Recurrent-based models

   In Recurrent-based models, there's no fixed window size concept during inference. These models process sequential data iteratively, adjusting their internal states dynamically based on the current input and previous states, allowing them to handle sequences of varying lengths effectively without explicit window size management.

   (iii) (5 points) Transformer-based models

   Transformer-based models window size during inference is typically handled through attention mechanisms. These models process input sequences in parallel, allowing them to handle sequences of arbitrary length. Attention mechanisms enable the model to attend to relevant parts of the input sequence, effectively capturing temporal dependencies without the need for explicit window size management, thus making them suitable for time series prediction tasks.

[1] K. Bandara, H. Hewamalage, Y.-H. Liu, Y. Kang, and C. Bergmeir, "Improving the accuracy of global forecasting models using time series data augmentation," *Pattern Recognition,* vol. 120, p. 108148, 2021.