National Tsing Hua University

1130IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: Ting, Kuan-Fu Student ID: 113003803

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.)

Answer:

Window size	Step	Best Val loss	Test MSE
10	15	170.6830	275.1708
10	5	8.5862	21.7506
30	15	151.9381	360.5837
30	5	6.1440	19.3879

Increasing the window size and decreasing the step size helps improve the model's performance.

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.

Answer:

(i)

Add 'Volume' without normalization.

Window size	Step	Best Val loss	Test MSE
30	5	541.9514	1149.7184

Add 'Volume' with normalization.

Window size	Step	Best Val loss	Test MSE
30	5	31.9623	36.6692

Adding 'Volume' without normalization will negatively affect the model's performance. If I add 'Volume' with normalization, it also negatively affects the model's performance. Therefore, 'Volume' may be a noise when predicting the stock price.

(ii) PCA analysis shows “Volume” is not helpful for prediction. Feature Importance Random Forest shows 'Open', 'Low', and 'Close' are the top 3 in importance. Feature Importance XGBoost shows 'High', 'Open', and 'Close' as the top 3 in importance. Therefore, the best input combination is the 'Open', 'High', 'Low', 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.)

Answer:

'Open', 'High', 'Low', and 'Close' with normalization.

Window size	Step	Best Val loss	Test MSE
30	5	486.4525	1151.3688

'Open', 'High', 'Low', and 'Close' without normalization.

Window size	Step	Best Val loss	Test MSE
30	5	19.3862	22.1860

If adding “Volume”, it needs to be normalized. If not, don’t use normalization for this case.

4. 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.

Answer:

Window size: How many time steps the model looks at per input.

Step size: How much we move the window forward each time.

Window size > Step size is most common.

“This lag parameter defines the input window size. It is typically combined with a stride (or step size) to generate overlapping input-output sequences.” (Elsworth, S., & Güttel, S. (2020). Time series forecasting using LSTM networks: A symbolic approach. *arXiv preprint arXiv:2003.05672*.)

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

Answer:

The transforms in the time domain are the most straightforward data augmentation methods for time series data. Most of them manipulate the original input time series directly, like injecting Gaussian noise or more complicated noise patterns such as spike, step-like trend, and slope-like trend. (Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2020). Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478*.)

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

(i) (5 points) Convolution-based models

(ii) (5 points) Recurrent-based models

(iii) (5 points) Transformer-based models

Answer:

Model Type	Variable Input Length	Recommended Inference Strategy	Match Training Window Size	Citation
Convolutional Neural Network (CNN)	No	Fixed-length sliding window inference	Yes	arXiv:1803.01271
Recurrent Neural Network (RNN/LSTM/GRU)	Yes (but with caution)	Fixed-length recent window + recursive forecasting	Recommended	arXiv:1409.2329
Transformer (Informer, Autoformer, etc.)	No	Fixed-length encoder input + autoregressive decoding	Yes	arXiv:2012.07436