National Tsing Hua University

11320IEEM 513600 Deep Learning and Industrial Applications

Homework 4

Name: 董少霖 Student ID: 113034567

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.)

Experiment	Window Sze	Step	MSE(Validation)
A	10	15	176.15
B	20	10	100.11
C	30	5	6.16

As the window size increased and step size decreased, the model's performance (in terms of MSE) improved. A larger window allows the model to capture more temporal dependencies, giving it better context to forecast the next value (while larger window size improves temporal context, it may also risk overfitting if training data becomes insufficient). Meanwhile, a smaller step size generates more overlapping windows, increasing the training sample size. However, this also increases redundancy and training time.

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.

   Add 'Volume' to the feature set and retrain the same LSTM model under identical settings. However, the validation MSE increased significantly — from **6.16** to **641.97**. This unexpected degradation suggests that the raw `Volume` values may have introduced scale imbalance or irrelevant noise, which disrupted the model's learning.

   (ii) (15 points) Explore and report on the best combination of input features that yields the best MSE. Briefly describe the reasons for your attempts and analyze the final, optimal input combination.
   The window size and step are the same as in Q1.

Close	Close, Open	C+Hign, Low	OHLC	OHLC+Volume
10.27	13.90	73.73	0.54	1063.23

The OHLC combination achieved the lowest MSE (0.54) while adding 'Volume' significantly degraded performance. These results highlight that **feature quality** and **data preprocessing** are more crucial than quantity alone in time-series prediction tasks.

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.)
After applying normalization, the validation loss decreased from **1063.23** (unnormalized) to **640.30**, indicating that normalization significantly improves training stability and prediction performance. This is because the Volume feature has a much larger scale compared to price features, causing scale imbalance that hinders model convergence. This observation aligns with Ioffe & Szegedy (2015)[1], who proposed that normalization helps reduce internal covariate shift, allowing neural networks to train faster and more reliably.

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 Lab 4, setting the window size smaller than the step size ensures that the input sequences do not overlap heavily, which prevents data redundancy and reduces training time. I think this is reasonable because excessive overlap can cause overfitting and make the model memorize patterns instead of generalizing.[2]

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 effective data augmentation method for time-series data is **time warping**. Time warping involves randomly stretching or compressing segments of the time series along the temporal axis without altering the overall pattern. This technique helps models become more robust to temporal variations and improves generalization. Time warping can simulate real-world variations in sequential data, especially for tasks like activity recognition and stock prediction. According to Um et al. (2017)[3], time warping

---

[1] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). pmlr.

[2] Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. In *Web-Age Information Management - 15th International Conference, WAIM 2014, Proceedings* (pp. 298-310).

[3] Um, T. T., Pfister, F. M., Pichler, D., Endo, S., Lang, M., Hirche, S., ... & Kulić, D. (2017, November). Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th ACM international conference on multimodal interaction* (pp. 216-220).

significantly improves the performance of deep learning models on physiological time-series datasets.

6. Discuss how to handle window size during inference in different model architectures (approximately 150 words):
   (i) (5 points) Convolution-based models
   Window size determines the receptive field of convolution layers. During inference, the input sequence length must match or exceed the window size used during training to ensure meaningful feature extraction.

   (ii) (5 points) Recurrent-based models
   RNNs, including LSTMs, naturally handle variable-length sequences. However, for consistent performance, it is recommended to use a similar window size during inference as in training, since RNNs are sensitive to sequence length.

   (iii) (5 points) Transformer-based models
   Transformers process sequences in parallel using self-attention. During inference, longer sequences can be handled if memory permits. Nonetheless, the model may perform best when the inference window size is close to the training configuration to preserve positional encoding patterns.