National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name: Li-Chiao Wang (王麗喬)   Student ID: 112034804

Completed on May 2, 2024

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.)

Here's a summary table showing MSE for different configurations:

Window Size	Step	MSE
10	15	415.55994
5	**10**	**136.61443**
8	15	353.13074
10	18	791.5147

The table indicates that the optimal configuration involves a window size of 5 and a step of 10. We observed that reducing the window size and step lowers the MSE. This suggests that employing smaller windows and steps allows the model to capture finer local patterns in the data, thereby improving prediction accuracy. On the other hand, keeping a fixed step while reducing the window size or increasing the step with a constant window size increases the MSE, showing that improper settings can worsen performance. To achieve the best results, it's crucial to balance window size, step, and model complexity.

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

Incorporating 'Volume' as an additional input feature led to a significant degradation in model performance, as indicated by the MSE increasing from 136.61443 to 1333.71. This suggests that including 'Volume' might have introduced noise or irrelevant information, adversely affecting the model's ability to accurately predict the target variable.

(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.

To find the best combination of input features, I experimented with different combinations of 'Open', 'High', 'Low', 'Close', and 'Volume'. The model with 'Open', 'High', and 'Low' features yielded the lowest MSE of 108.073906. This suggests that these features likely contain the most predictive information for the model, likely capturing key aspects of price movements and volatility.

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.

In our experiment, we observed a significant enhancement in the model's performance through input normalization. Without normalization, the MSE was notably higher at 136.61443. However, upon normalizing inputs, the MSE dropped drastically to 8.7246626e-05, indicating a substantially improved fit of the model to the data. This underscores the importance of normalization in accelerating convergence and attaining superior results. Normalization likely facilitates faster convergence and better outcomes by scaling inputs to a comparable range, thereby aiding optimization and preventing any single feature from overpowering the learning process.

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.

Typically, the window size should not be smaller than the step size to prevent data point overlap in consecutive windows. However, for Lab 4, the training data is abundant, eliminating the need for a window size smaller than the step. This avoids redundant information in model training, potentially enhancing the model's stability and predictive performance over time.

Reference: Rolling-Window Analysis of Time-Series Models (https://se.mathworks.com/help/econ/rolling-window-estimation-of-state-space-models.html)

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

Dynamic Time Warping (DTW) is a potential technique for measuring similarity between two sequences, often used in time-series analysis. To augment time-series data using DTW, one method involves applying time warping to stretch or compress segments of the series while maintaining its overall shape. This involves randomly selecting portions of the time series and applying DTW to align them with similar segments within the same series or across different series. By introducing slight variations in the temporal alignment, this method generates diverse augmented samples while preserving the underlying patterns in the data.

Reference: X. Yang, Z. Zhang, X. Cui and R. Cui, "A Time Series Data Augmentation Method Based on Dynamic Time Warping," 2021 International Conference on Computer Communication and Artificial Intelligence (CCAI), Guangzhou, China, 2021, pp. 116-120, doi: 10.1109/CCAI50917.2021.9447507.

6. Discuss how to handle window size during inference in different model architectures:

(i) (5 points) Convolution-based models

Convolution-based models, like CNNs, use convolutional filters whose kernel size determines the window size. During inference, input images are adjusted to match a predefined window size. Techniques like zero-padding or reflection padding handle edge cases. For larger images, sliding window techniques can be employed.

(ii) (5 points) Recurrent-based models

Recurrent-based models, such as RNNs and LSTMs, determine the window size by the length of the input sequence. During inference, these models handle input sequences of different lengths by either truncating or padding them to a fixed length, ensuring consistent input size for processing.

(iii) (5 points) Transformer-based models

Transformers, known for their attention mechanism, efficiently process variable-length sequences. During inference, input sequences are directly processed without padding, enabling focus on relevant parts regardless of length. To streamline computation, sequences may be batched by length with padding applied within each batch.