

**National Tsing Hua University**  
**11220IEEM 513600**  
**Deep Learning and Industrial Applications**  
**Homework 4**

Name: 吳柏彥

Student ID: 112034525

**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 size	Train loss	Val loss
10	10	64.0178	1.7776
10	15	76.5054	8.5318
10	20	278.64444	1.6055
15	10	35.2167	288.9859
15	15	133.3476	202.3939
15	20	80.7348	176.1389
20	10	62.8532	71.8531
20	15	106.8191	66.8566
20	20	154.9112	127.5684

從上方的數據來看，window size 跟 step size 的大小關係，好像對於 loss 沒有一個很明顯的相關性，只能觀察出 step size 越大時會導致 train loss 提升，val loss 降低。

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.

Exclude volume Epoch : 100 Train loss : 76.5054 Best Val loss : 8.8101	Include volume Epoch : 100 Train loss : 814.4247 Best Val loss : 616.1341
---	--

從上方的數據可以發現，加入了 volume 之後會使 Train loss 跟 Val loss 都變大許多，可能是 volume 在資料裡面太大以及資料的值很分散。

- (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.

在我的嘗試中，我發現將所有特徵導入模型會有最好的表現，可以有更多的數據去讓模型更有效的預測結果，但也要看要預測什麼去調整要導入的參數。

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.)

With normalized(standard scalar)	Without normalized
Epoch : 100	Epoch : 100
Train loss : 0.0048	Train loss : 786.5021
Val loss : 0.0038	Val loss : 769.7033
Best val loss : 0.0024	Best val loss : 764.7719

從上方數據可以發現，透過 Normalized 可以使模型的效果優化許多，應該是史資料統一正規化之後，就比較不會有極值，就不會使模型有許多的離群值去干擾到模型的建構。

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.)

我認為並不強迫 window size 要小於 step size，從第一題來看可以明顯看到，在不同的 window size 配到不同的 step size 時都會有不同的結果，我覺得這兩個參數在模型當中是互相獨立的。

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

這種方法通過將時間序列向前或向後移動若干時間步來創建新的樣本。這種平移操作可以引入不同的時間偏移，從而生成具有不同相位或時間延遲的數據。例如，可以將時間序列向前或向後滾動一個或多個時間步，然後使用這些平移後的序列作為新的訓練樣本。這種方法可以增加數據的多樣性，提升模型的泛化能力。

參考文獻：

R. Kajdanowicz 等人，"Data augmentation for time series classification using convolutional neural networks," Expert Systems with Applications, vol. 140, 2020, pp. 112885.

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

(i) (5 points) Convolution-based models

用於序列預測任務時，可以通過將固定大小的卷積過濾器在輸入序列上滑動來處理窗口大小。這種滑動操作使模型能夠在每個步驟處理輸入序列的不同部分，保持固定的窗口大小並有效地捕獲局部模式。

(ii) (5 points) Recurrent-based models

推論期間窗口大小的處理更加動態。在訓練期間，模型逐步處理序列，保持內部狀態並適應不同的序列長度。模型可以接受不同長度的輸入序列，並使用其循環結構逐步處理這些序列，在每個步驟基於當前和先前的輸入上下文進行預測。

(iii) (5 points) Transformer-based models

在訓練期間，模型通過基於位置嵌入計算的注意力權重來學習關注輸入序列的相關部分。在推論期間，窗口大小隱式地通過輸入標記的序列長度管理。模型可以處理不同長度的輸入序列並生成預測，通過同時關注序列中的所有標記，利用自注意力機制在整個序列中捕獲依賴關係，而無需依賴固定的窗口大小。