

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: 5, Step: 5] -> Test MSE: 14.432878  
[Window: 5, Step: 10] -> Test MSE: 135.371918  
[Window: 10, Step: 5] -> Test MSE: 19.545347
```

從實驗結果可觀察到，步長過大會導致序列之間重疊減少，資料利用率下降，使模型難以學習連續性的特徵，進而提高誤差。而窗口大小過大時可能引入噪音或冗餘資訊，導致預測性能下降。綜合來看，找到步長與窗口間的平衡點，有助於提升時間序列模型的預測效果。

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.

```
[Window: 5, Step: 5] -> Test MSE: 1093.577515  
[Window: 5, Step: 10] -> Test MSE: 1366.181396  
[Window: 10, Step: 5] -> Test MSE: 1096.106934
```

將 Volume 作為額外特徵加入模型後，整體 MSE 非常高，未出現顯著的效能提升，甚至略高於未加 Volume 的情況。這反映出成交量雖是反映市場熱度的重要指標，但其與價格的關係較為間接，可能對模型造成額外噪音。

- (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.

```
[Window: 5, Step: 5] -> Test MSE: 16.386007  
[Window: 5, Step: 4] -> Test MSE: 10.928969  
[Window: 5, Step: 2] -> Test MSE: 3.088748
```

由第一題結果可知，在適當範圍內，較小的步長有助於提升模型表現。步長減少代表序列樣本間的重疊增多，讓模型在訓練時能見到更多重複的時間特徵，增強對序列模式的掌握。因此，(window = 5, step = 2) 時的 MSE 最低，顯示細緻的資料滑動能有效提升預測精度。

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.)

```
[Window: 5, Step: 5] -> Test MSE: 13.876109
[Window: 5, Step: 4] -> Test MSE: 12.977309
[Window: 5, Step: 2] -> Test MSE: 2.607472
```

我使用了 `MinMaxScaler` 對資料進行標準化，將數據縮放到 0~1 範圍，觀察結果顯示標準化後在 (window = 5, step = 5) 與 (window = 5, step = 2) 時均有輕微進步 (MSE 降低)。這是因為標準化能消除不同特徵間的尺度差異，使模型更容易收斂，學習速度加快。然而在 (window = 5, step = 4) 的組合中，反而使 MSE 上升。這可能與步長 4 導致的資料覆蓋度及重疊度變化有關，模型面臨的訓練樣本分布恰好較難捕捉其模式，使模型在此組合下出現過擬合現象。整體來說，標準化大多有助於提升預測效果，但對部分參數組合仍需額外調整。

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.)

在時間序列預測中，步長決定了滑動窗口在資料上移動的間隔，而窗口大小則決定了每次輸入模型的時間步數。通常建議步長應小於窗口大小，這樣可以確保資料樣本之間有足夠的重疊，從而捕捉時間序列中的連續性和模式。因此我認為問題敘述的是錯誤的。

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

在時間序列資料增強中，**Jittering**（抖動）是一種常用且有效的方法，透過在原始資料中加入隨機噪音來模擬現實世界中的微小變化。這種技術有助於提升模型對於資料中微小變動的穩健性，並減少過擬合的風險。

Jittering 通常是將服從高斯分佈的隨機噪聲加入到每個時間點的資料中，從而生成新的樣本。這種方法特別適用於感測器資料等本身就含有噪聲的時間序列資料。

[參考資料](#)

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

- (i) (5 points) Convolution-based models

在卷積模型中，窗口大小通常依賴 `filter` 的大小及模型架構設計。若輸入序列很長，可將序列拆解為與訓練時相同長度的片段（窗口），逐步滑動這些片段送入模型，並將多次預測結果結合（如取移動平均）以獲得完

整序列的預測。此外，因為卷積是局部操作，即使輸入長度變化，卷積層仍可自動處理，但為了與訓練一致，建議使用固定大小的窗口。

另一種方法是使用全序列推論，將整段序列直接輸入，讓模型在卷積過程中自行提取局部特徵。

(ii) (5 points) Recurrent-based models

在遞迴模型中，例如 LSTM 或 GRU，推論時的窗口大小通常與訓練時保持一致，以確保時間依賴特徵一致。推論時可使用固定長度的滑動窗口，從序列中逐步截取片段進行預測，特別是在長序列中採用移動窗口的方式，逐步推進一段一段預測。

另一種方法是序列遞迴推理，模型每次只輸入最新的資料點，同時保留上一步的隱藏狀態，實現連續預測，這種方法適合即時序列預測場景。

(iii) (5 points) Transformer-based models

在 Transformer-based 模型中，由於 Transformer 本身不具備內建的時間步限制，因此可直接將整段序列輸入模型進行一次性預測，適用於短序列情境。對於長序列，常用方法是將序列切分為與訓練時相同的固定長度的窗口，進行滑動預測，或採用重疊窗口以捕捉跨段資訊。此外，對於自回歸型 Transformer，可實施動態窗口擴展，每次將前一步預測結果拼入序列中，逐步向後預測。