

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	Test MSE
15	3	0.001433
30	5	0.086222
60	10	0.003356

在本次實驗中，分別測試了三組不同的視窗大小與步長組合。當 Window Size 為 15、Step Size 為 3 時，模型達到了最低的 Test MSE (0.001433)，預測表現最佳。當 Window Size 增加到 30、Step Size 為 5 時，Test MSE 顯著上升至 0.086222，顯示模型準確度下降。進一步將 Window Size 擴大到 60、Step Size 為 10 後，Test MSE 又下降至 0.003356。結果顯示，適當增加視窗大小有助於捕捉長期趨勢，提升模型穩定性與預測效果。

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.
 - (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.

- (i) MSE with Volume added: 0.010984

在將 Volume 作為額外特徵後，模型的 Test MSE 為 0.010984，相比僅使用 Close 特徵時的 MSE 有所下降。這顯示 Volume 這個特徵能提供額外的市場活動資訊，提升模型的預測能力。

Feature Set	Test MSE
All features	0.000440
Close + Volume	0.001152
Open + High + Low	0.001327
OHL + Close	0.004377
Close only	0.007942

然而，單獨加入 Volume 並未帶來最佳表現，說明 Volume 雖有幫助，但其效果有限，仍需搭配其他特徵以進一步優化模型準確度。從實驗結果可見，使用所有特徵 (All features) 時，Test MSE 最低 (0.000440)，表現最佳。這代表在本資料集中，不同特徵 (Open, High, Low, Close, Volume) 能夠互補，提供更多樣且完整的資訊，幫助模型更準確地預測未來趨勢。單獨使用 Close 或少部分特徵時，模型表現明顯下降，證明多元特徵結合對模型準確度具有重要貢獻。因此，最終選定使用全特徵作為最佳輸入組合。

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.)

Normalized MSE: 0.007237

Raw Input MSE: 5444.394043

在本次實驗中，將輸入資料進行正規化（Normalization）處理後，模型的測試 MSE 大幅下降至 0.007237，相比於未正規化時高達 5444.394043 的 MSE，有極為顯著的改善。這明確顯示正規化對於模型性能提升的效果。主要原因是，正規化能將各特徵縮放至相近的數值範圍，降低特徵間尺度差異，從而避免因特徵值過大或過小導致的梯度消失（vanishing gradients）或梯度爆炸

（exploding gradients）問題。此外，輸入分布的一致性也有助於加速模型的收斂過程，提升訓練穩定性與最終的泛化能力。根據 Goodfellow, Bengio, 與 Courville (2016) 在《Deep Learning》一書中的說明，適當的特徵縮放是深度學習訓練中不可或缺的重要步驟，能顯著影響模型的效能與收斂速度。

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.)

在時間序列資料處理中，設定 window size 小於 step size，可以有效減少資料間的重疊程度。過多的重疊可能導致模型過度記憶訓練資料（overfitting），因為相鄰的樣本過於相似，模型容易學到無意義的細節而非真正的資料趨勢。相反地，減少重疊使樣本間具有更多變異性，有助於提升模型的泛化能力，特別是在預測未來變化趨勢時能更靈活地適應資料的自然波動。根據 Brownlee (2017) 的說明，適當控制重疊程度是時間序列建模中避免過擬合、強化未來預測能力的重要策略之一。

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

在時間序列資料中，一種常用且有效的資料增強（Data Augmentation）方法是「時間扭曲（Time Warping）」。此方法透過對時間軸進行非線性擴展或壓縮，使資料在不同區段產生輕微的速度變化，從而製造出新的變異樣本。藉由改變資料的時間進程，但保留整體結構與趨勢特性，時間扭曲能增加資料集的多樣性，有助於提升模型對變異模式的辨識與適應能力。此技術特別適合用於時間序列的分類與預測任務，能有效緩解因訓練資料不足導致的過擬合問題。根據

Um et al. (2017) 的研究，時間扭曲等增強方法能顯著提升穿戴式感測器資料在帕金森氏症監測任務中的模型效能，驗證了其在時間序列領域的實用性與有效性。

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

(i) Convolution-based models

在推論階段，卷積型模型（如 CNN）通常要求輸入的 window size 與訓練時保持一致，因為卷積核（filter）的感受也是固定設計的。若推論輸入與訓練時不同，常透過**裁切（trimming）**或**補零（zero padding）**，以調整序列長度至預期尺寸。此外，若使用全卷積網路（Fully Convolutional Networks），有時也允許輸入不同長度，但必須注意輸出層尺寸的對齊問題。

(ii) Recurrent-based models

遞迴型模型（如 RNN、LSTM、GRU）能靈活接受變長輸入，推論時不需嚴格固定 window size。不過，過長的序列可能引發記憶遺失（vanishing gradients）或延遲推論時間，因此常設定最大輸入長度，或將長序列分割成多個固定小視窗（sub-sequences）逐步推論，以平衡效能與準確度。

(iii) Transformer-based models

Transformer 模型理論上支持可變長輸入，但因 self-attention 計算成本隨輸入長度平方成長，推論過長序列時會大幅增加記憶體與時間消耗。實務上常設定固定 window size，並結合**滑動視窗（sliding window attention）**、稀疏注意力（sparse attention）或分段處理（chunking）技術，以支援超長序列推論。