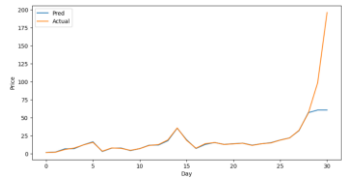
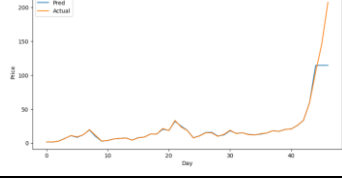
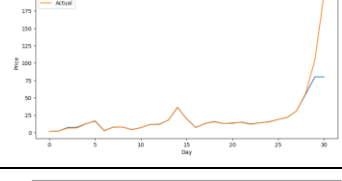
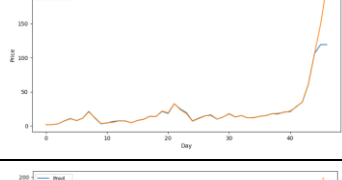
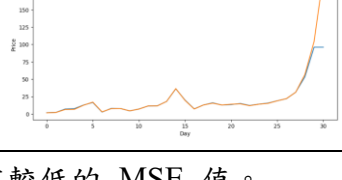


113034553_黃熾云_HW4

- 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.

Experiments	Window sizes	Steps	MSE	Figure
Original	10	15	639.2719	
1	15	10	208.2293	
2	8	10	473.1111	
3	20	8	181.1532	
4	8	15	342.8915	

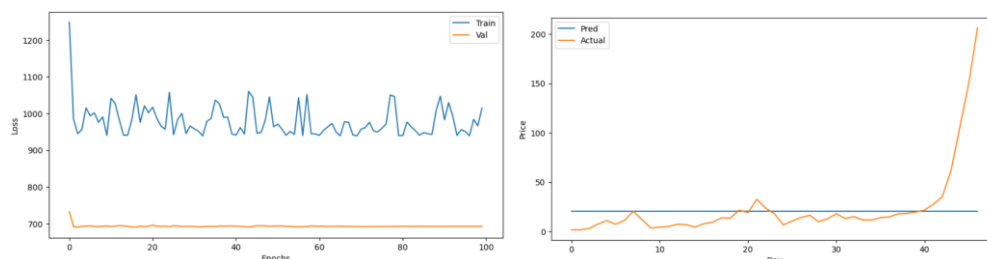
根據大致觀察，window size 大、step size 小會有較低的 MSE 值。

- Include 'Volume' as an additional input feature in your model. Discuss the impact of incorporating 'Volume' on the model's performance.
 - 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.

在模型中加入 Volume 作為額外特徵後，模型的 MSE 明顯上升(由原本的約 181.1532 增加到 1379.2711)，代表 Volume 並沒有提供有用的額外資訊，反而引入了更多雜訊，降低了預測表現。可能原因包括：

- Volume 數據波動大且噪聲多，難以捕捉穩定的預測模式。
- Volume 與價格變化的關聯性不強，模型無法有效利用這個特徵。

- 在目前的簡單模型架構下，無法深入理解 Volume 與價格之間的隱含關係。因此，加入 Volume 對本次模型表現有負面影響。



(ii) 最佳特徵組合探索透過多次嘗試，發現根據不同特徵組合的實驗結果，同時使用 ['Open', 'High', 'Low', 'Close'] 四個特徵時，模型表現最佳，MSE 為 181.1532，顯著優於其他組合。具體分析如下：

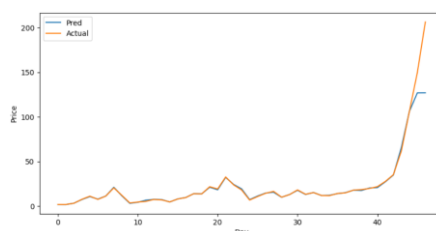
- 單獨使用 ['Close'] 預測 (MSE 289.0609) 效果不如同時考慮其他價格資訊。
- 僅使用 ['Open', 'Close'] (MSE 197.7160) 有一定改善，但仍不如完整四個特徵。
- 單獨使用 ['High'] 或 ['High', 'Low'] 效果較差 (MSE 分別為 432.7199 與 239.7190)。

因此，整合 Open、High、Low、Close 四個價格指標，可以讓模型獲得更全面的市場資訊，有效降低預測誤差。相比之下，如第一小題的實驗，加入其他特徵（如 Volume）反而會增加資料的雜訊與不確定性，影響預測準確率。

Feature	MSE
['Open', 'High', 'Low', 'Close']	181.1532
['High', 'Low']	239.7190
['Open', 'Close']	197.7160
['Close']	289.0609
['High']	432.7199

3. 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

使用正規化 (Normalization) 處理資料後，模型訓練的收斂速度加快，且最終測試 MSE 有明顯下降。正規化可以將不同量級的特徵縮放到相近範圍（如 0~1 或標準常態分布），避免特徵之間因尺度不同而影響模型學習效率。使用 MinMaxScaler 後達到 MSE 145.8969。



此外，正規化能穩定梯度下降過程，避免模型訓練過程中出現梯度爆炸或消失的問題。

根據文獻【Cooijmans, T., Ballas, N., Laurent, C., Gülçehre, Ç., & Courville, A. (2016). Recurrent batch normalization. arXiv. <https://arxiv.org/abs/1603.09025>】，正規化是提升深度學習模型效能的常見且有效方法。本實驗亦觀察到，使用正規化能有效降低最終 MSE，證實其正面影響。

4. 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.

實驗顯示，當 window size 大於或等於 step size 時，模型的 MSE 整體表現更佳。例如，當 window size = 20 且 step size = 8 時，模型達到最低 MSE (181.1532)。猜測為重疊更多的資料片段可以讓模型看到更多上下文，增加每個時間點的訓練樣本次數，有助於提升泛化能力。時間序列資料往往連續性高，使用較大的 window size 可以提供更完整的趨勢結構資訊。多個高度重疊的片段能幫助模型更穩定地學習相似模式，避免過度依賴個別資料點。【[Wang et al., 2021, A New Multi-Scale Sliding Window LSTM Framework \(MSSW-LSTM\): A Case Study for GNSS Time-Series Prediction](#)】。

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

一種適合時間序列資料的資料增強方法是時間扭曲 (Time Warping)。時間扭曲指的是隨機地將資料段在時間軸上進行拉伸或壓縮，模擬現實中事件發生快慢的不同，有助於增加資料多樣性並提升模型的泛化能力。該方法利用 Dynamic Time Warping (DTW) 和 shapeDTW 的對齊特性，將樣本模式的特徵扭曲以匹配參考模式的時間步驟，從而生成新的訓練樣本。參考資料：【Iwana, B. K., & Uchida, S. (2020). Time series data augmentation for neural networks by time warping with a discriminative teacher. arXiv. <https://arxiv.org/abs/2004.08780>】。

6. Discuss how to handle window size during inference in different model architectures: (i) (5 points) Convolution-based models (ii) (5 points) Recurrent-based models (iii) (5 points) Transformer-based models.

(i) Convolution-based models

卷積網路通常使用與訓練時相同的固定窗口大小 (如 10 或 20 步)，利用局部感知能力提取時間內的關鍵特徵。維持固定大小有助於保持特徵學習的一致性。

(ii) Recurrent-based models

RNN、LSTM 等模型可以接受變長序列，但為了保持運算效率，通常仍會設定固定的時間窗口長度，並使用 padding 或 masking 技術處理輸入長度的差異。

(iii) Transformer-based models

Transformer 架構理論上能處理任意長度的序列，但由於注意力機制的運算複雜度為 $O(n^2)$ ，實際運算時會限制最大序列長度 (例如取最近的 256 步)，或使用 sliding window attention 等技術降低推理成本。