

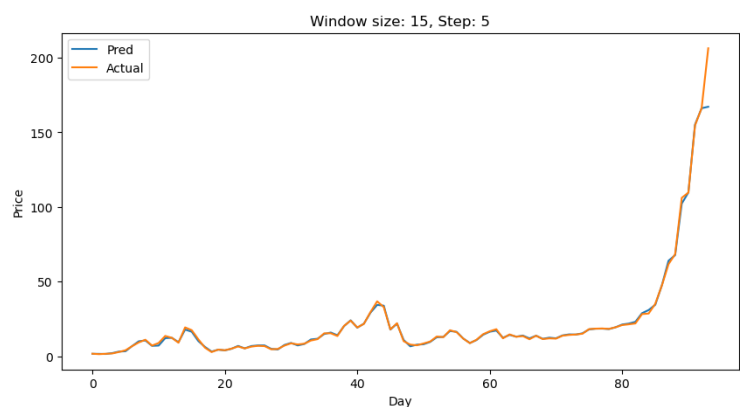
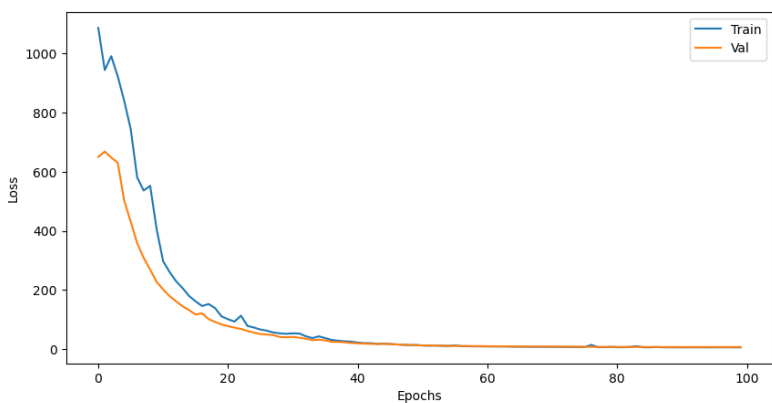
Homework 4

學號：113034510 姓名：蘇懂

1. 以下實驗結果為 window sizes 和 steps 的不同數值選擇。經不同組合嘗試，發現 steps=5 有明顯較佳的效果，因此持續嘗試其餘 window sizes：

Window Sizes	Steps	Train Loss	Val Loss	Best Val Loss
5	5	5.234203	7.690431	7.41231
5	10	25.74013	124.4316	124.4316
5	15	76.86901	187.9526	187.9526
5	20	134.1445	403.5409	403.5409
5	30	150.9401	376.3175	376.3175
10	5	5.232036	8.39469	8.278263
15	5	6.081079	7.308459	7.186388
20	5	6.905857	8.476655	8.476655
30	5	7.35225	8.907722	8.904067

可以發現 steps 越大，Val Loss 表現越差，而 window sizes 的影響反而沒那麼劇烈。最佳的實驗結果為 window sizes=15、steps=5，以下是最佳的視覺化圖表：



2. (i). 實驗如下表，特徵加入 Volume 後，使用 window sizes=15、steps=5，比較可以發現模型的 Train loss、Val loss、Best Val loss 都顯著增加，表現下降。主因可能是因為 Volume 的 Scale 相較於其他特徵有很大的範圍，造成模型的不穩定。因此，適當的前處理

很重要，例如 Normalization 或其他特徵處理，再去做特徵選擇，才能最大化模型成果。

(ii). 實驗如下表，額外嘗試單特徵的效果，評估其 Val Loss 後，再逐一加入較好的來做觀察，發現原來的選擇還是較佳，猜測可能是 Open、High、Low、Close 本身就有一定的特徵貢獻與相關性，所以都使用才有較好結果。

Features	Train Loss	Val Loss	Best Val Loss
['Open', 'High', 'Low', 'Close']	6.081079	7.308459	7.186388
['Open', 'High', 'Low', 'Close', 'Volume']	1062.355	658.3608	649.3255
['Open']	11.02335	12.31343	12.31343
['High']	11.05811	12.66242	12.66242
['Low']	11.38635	13.30437	13.30417
['Close']	11.24855	12.38765	12.38765
['Open', 'Close']	6.855598	8.125357	8.125357
['Open', 'High', 'Close']	8.812822	10.25589	10.25589

- 對輸入資料進行 Normalization 後，模型表現明顯改善，MSE 大幅減少，尤其是比較有使用 Volume 的情況，應證了第二題的猜想。這是因為神經網路對不同尺度的數據敏感，Normalization 能加快收斂速度並減少梯度爆炸、消失問題。根據文獻 (Ioffe & Szegedy, 2015)，Normalization 在深度學習中是提升穩定性與精度的關鍵步驟。

Features	Normalized	Train Loss	Val Loss	Best Val Loss
['Open', 'High', 'Low', 'Close']	False	6.081079	7.308459	7.186388
['Open', 'High', 'Low', 'Close', 'Volume']	False	1062.355	658.3608	649.3255
['Open', 'High', 'Low', 'Close']	True	0.001636	0.001472	0.001304
['Open', 'High', 'Low', 'Close', 'Volume']	True	0.001535	0.00161	0.001398

(Reference : Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML) (pp. 448–456). <https://arxiv.org/abs/1502.03167>)

- 對於要求 window size 小於 step size 猜測可能是為了避免資料

重疊，減少 overfitting 風險。但實際上我認為不一定正確，也如以上實驗結果所示，較小的 step size 能提供更密集的樣本，有助於模型學習時間相關性。文獻 (Zhao et al., 2017) 也指出，適當的重疊有助於提升序列模型效能。

(Reference : Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM Network: A Deep Learning Approach for Short-term Traffic Forecasting. IET Intelligent Transport Systems, 11(2), 68–75. <https://doi.org/10.1049/iet-its.2016.0208>)

5. 有一種資料增強方法為 Time Warping，透過非線性地拉伸或壓縮資料的時間軸，使模型能學會在不同速度下的序列特徵，提升對時間變異的 Robustness。根據文獻 (Um et al., 2017)，這能有效提升模型在感測器數據分類任務中的泛化能力，常與其他增強方法如 Jittering 或 Window Slicing 一同使用。

(Reference : Um, T. T., Pfister, F. M. J., Pichler, D., Endo, S., Lang, M., Hirche, S., Fietzek, U., & Kulić, D. (2017). Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks. Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI), 216–220. <https://doi.org/10.1145/3136755.3136817>)

6. (i). Convolution-based：通常採用固定大小的滑動窗口進行卷積操作。在預測時，需保留與訓練相同的窗口大小以保證特徵對齊。一種策略是使用滑動窗口方式逐步推進，對整段序列分段預測。
(ii). Recurrent-based：適合處理可變長度序列。在預測時可使用與訓練相同的窗口長度，或將整段序列分段送入模型，使用狀態儲存機制持續推進，無需嚴格限制 window size。
(iii). Transformer-based：通常受限於計算資源，實務中常使用固定長度輸入。常使用 sliding window 或 cache-based incremental decoding 技術維持記憶，處理長序列資料時可動態調整窗口大小，並保留先前注意力資訊。