

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	steps	Best Val loss
1(original)	10	15	111.111
2	10	20	721.4302
3	15	10	7.6033
4	15	1	1.9887

如圖可以發現當 step size 大於 window size 時，MSE 會很大，表示當跳過中間 (step size-window size) 可能會導致數據不足。另外當設 step size 為 1 時可以有最小的 MSE，可發現當 step size 較小時，可以有較多數據來 train model

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. 加入“Volumn”後，發現到 train 不起來，可以觀察到“Volumn”的數值與其他 feature 的範圍相比大約多了 $1e+7$ 倍，因此推測可能是在做 backpropagation 時梯度爆炸所導致，能解決此現象的方法可以做標準化或 min-max。
- ii. 當 step size 設為 1，window size 設為 15 時，MSEloss 最小。Step size 較小時可以有較多資料來 train model，資料樣本為(總樣本-window size)，window size 則一次取半個月的資料預測下一筆，主要是因為這個 data 看起來沒有比較長期季節性趨勢。

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.)

Label 不 normalization(min-max)(MSE 也會被 normalize)

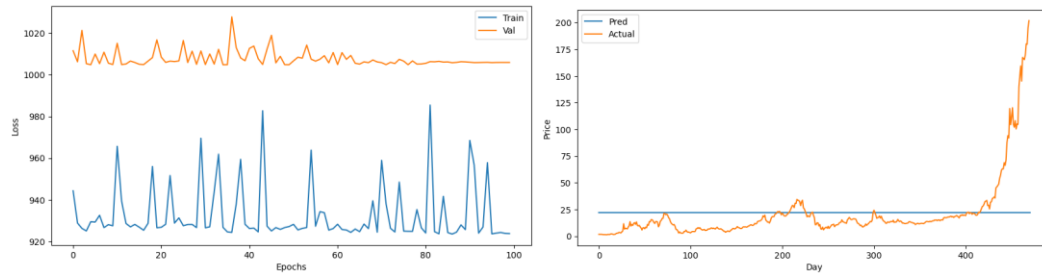
加入 normalization(min-max)後，可以發現 train 不起來，推測是因為特徵數不多或特徵單位一致(數值約 1-220)，以下是實驗結果:

- 數值範圍

最小值:		最大值:	
Open	1.2989	Open	213.08
High	1.3175	High	218.67
Low	1.2339	Low	211.63
Close	1.2619	Close	216.14
dtype:	float64	dtype:	float64

● 實驗

當 step size 設為 1，window size 設為 15 時



Reference:針對 label 是否需標準化

<https://stats.stackexchange.com/questions/111467/is-it-necessary-to-scale-the-target-value-in-addition-to-scaling-features-for-re>

<https://stackoverflow.com/questions/36540745/pre-processing-data-normalizing-data-labels-in-regression>

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.)

我認為應該 window size 應該要大於 step size。Window size 較大，可以捕獲更多的歷史訊息，提高模型準確率。

Reference：Time-series forecasting with deep learning: a survey

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

依 time-domain 來看，分為以下幾個 basic methods

- Slicing：將長度不一的時間序列，切成等長 slice
- Warping：針對某一時間段做 upsampling(2 個點變成 1 個，可使解析度降低時間縮短(使用 slicing))或 downsampling
- Flipping：正負交換，但針對 ECG 資料會有問題
- Noise injection：加入噪音
- Label expansion：label 附近都是對的(ex:邊界有不確定存在)

Reference: Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2021). Time

Series Data Augmentation for Deep Learning: A Survey

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. Window(卷積核)大小通常跟模型架構有關，在 inference 時，會使用 padding 和 stride 來影響有效 window 大小，透過期望感受野大小決定 pooling 的 window size
- ii. Window size：時間序列的長度
在 inference 時，若有不同長度的輸入可以使用 masking 的方式(補 0)或直接以最短序列者為長度截斷其他較長序列
- iii. 透過 attention 機制看 features 彼此關聯度，較長的序列可以捕捉長距離依賴性。Multihead 的機制可以關注此序列的不同面向