National Tsing Hua University

Deep Learning and Industrial Applications - Homework4

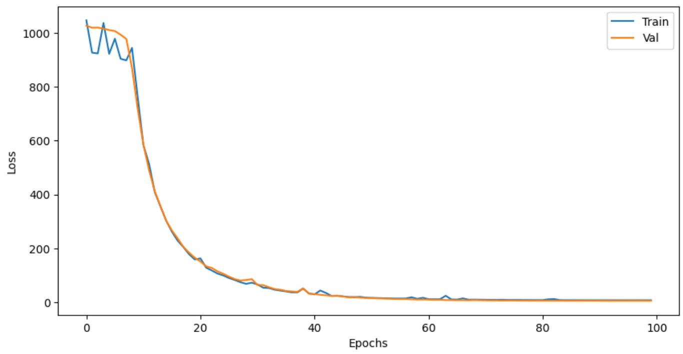
Name: 謝伊婷 Student ID: 112034533

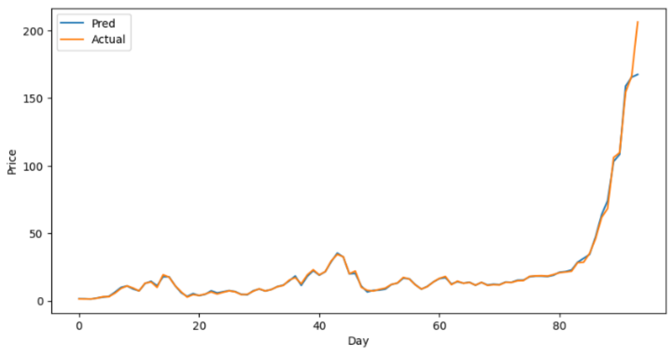
1.

id	window_size	step	Train loss	Val loss	Best Val loss
1	10	15	92.791	227.271	227.271
2	10	5	5.371	11.998	11.995
3	20	5	6.753	12.188	12.153
4	20	30	460.759	0.287	0.280
5	15	10	45.145	52.625	52.625
6	5	10	45.558	54.117	54.117
7	10	5	7.373	5.375	5.322

實驗結果顯示，較小的 step 可能有助於提高預測的準確性，但需要避免 step 過小而導致過擬合。而 window_size 為 20 和 step 為 30 的組合，訓練損失非常高、驗證損失非常低，可能是過擬合。

這些結果中，我認為表現最好的是 window_size 為 10 和 step 為 5 的這組，顯示了相對較低的 Train loss 和 Val loss，表示這組設定在預測股票價格上的表現較好 (下圖是這組參數下 model performance 的 Visualizing)。





2. (i)

id	window_size	step	Train loss	Val loss	Best Val loss	features
7	10	5	7.373	5.375	5.322	'Open', 'High', 'Low', 'Close'
8	10	5	911.313	1507.868	1497.500	'Open', 'High', 'Low', 'Close', 'Volume'

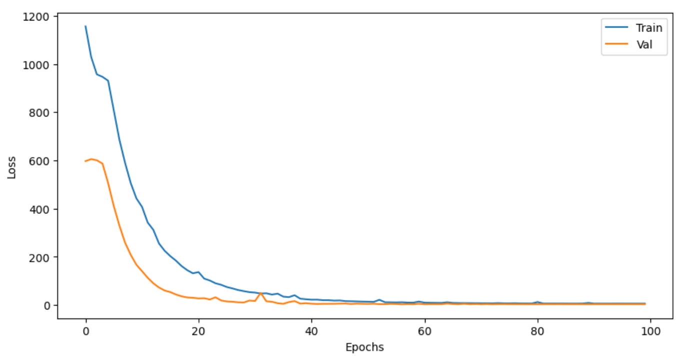
參數延續使用表現較好的 window_size 和 step，特徵加入「Volume」後，模型的 Train loss、Val loss 以及 Best Val loss 都顯著增加，模型表現下降。可能是因為「Volume」引入了額外的 noise 或者 data 維度的增加導致模型出現過擬合。因此，雖然成交量可以反映市場的活躍程度和可能的價格壓力，但在這個 case 中，它似乎沒有幫助模型更準確地預測股價，反而增加了模型的預測錯誤。這表示了，在加入新的 features 時，需要對 features 仔細評估和選擇。

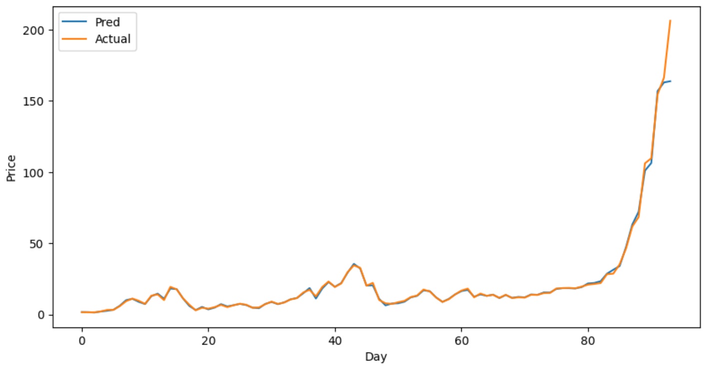
(ii)

id	window_size	step	Train loss	Val loss	Best Val loss	features
7	10	5	7.373	5.375	5.322	'Open', 'High', 'Low', 'Close'
8	10	5	911.313	1507.868	1497.500	'Open', 'High', 'Low', 'Close', 'Volume'
9	10	5	11.578	19.173	19.149	'High'
10	10	5	5.812	3.107	2.819	'Open', 'High', 'Low', 'Close'

參數延續使用表現較好的 window_size 和 step，嘗試了三種不同的 features 組合，在沒有 normalized inputs 情況下，我認為 best combination of input features that yields the best MSE 是使用'Open', 'High', 'Low', 'Close'，可以使模型達到最低的 MSE，

我認為可能因為這些 features 反映市場的價格行為，而「Volume」雖然反映了交易量，但在預測未來價格上可能會有額外的 noise。因此，best combination of input features that yields the best MSE 是「Open, High, Low, Close」。(下圖是上表中紅字這組參數下 model performance 的 Visualizing)。





3.

id	Window size	step	Train loss	Val loss	Best Val loss	features	normalized inputs
11	10	5	1033.876	723.982	712.817	'Open', 'High', 'Low', 'Close', 'Volume'	without
12	10	5	0.001	0.003	0.003	'Open', 'High', 'Low', 'Close', 'Volume'	with
13	10	5	8.623	16.589	16.570	'Open', 'High', 'Low', 'Close',	without
14	10	5	0.002	0.001	0.001	'Open', 'High', 'Low', 'Close',	with

從實驗結果來看，normalized inputs 對 model 的表現有顯著的正面影響。

比較 id 為 11 和 12 的結果，當加入標準化處理後，訓練損失、驗證損失和最佳驗證損失都從千位數左右大幅降低到接近零的水平。同樣地，比較 id 為 13 和 14 的結果，也有類似的趨勢，其中標準化後的 loss 明顯低於未標準化的 loss。

這些結果表示，feature normalization 能幫助 model 更好地理解和學習數據，在這個 case 中，有效地增加模型的預測準確性。而下面這篇被收錄在 JMLR 的 paper 也可以支持這個論點。

https://proceedings.mlr.press/v37/ioffe15.html

4.

我認為 window size 不應該小於 step size，因為 window size 用來定義每個 sequence 中要包含多少時間點的 data (例如 window size 設為 10 表示每個 sequence 都包含 10 天的股價)，而 step size 則定義了從一個 sequence 到下一個 sequence 要跳多少筆的資料 (例如步長設為 15，序列每次會跳 15 天，每個序列起始點是從第 1 天、第 16 天、第 31 天...)，這樣有一些 data 被跳過，可能會遺失一些重要資訊。

理想的設定是窗口大小等於或大於步長，特別是在時間序列分析中，連續性是非常重要的。

5.

One effective method for data augmentation in time-series data involves jittering, scaling, and rotating the data. This technique, discussed by Um et al. (2017), generates new variations from existing sequences, enhancing model robustness by simulating potential real-world variations. These transformations help improve the model’s ability to generalize from training data to unseen data, which is crucial in scenarios with limited data samples. This method is particularly applicable to wearable sensor data, where capturing diverse motion patterns is essential for accurate monitoring and prediction.

Reference: https://arxiv.org/abs/1706.00527 & google search

6.

(i) **Convolution-based models**:

- Window size is fixed during training to define kernel size or receptive field.
- Same window size must be maintained during inference to ensure accurate data interpretation.
- Essential for models as it influences how input data's spatial or temporal relationships are understood.

(ii) **Recurrent-based models**:

- Window size during training defines sequence length or the number of time steps for backward look to make predictions.
- Window size can be flexible during inference but is often kept consistent with training for performance optimization.
- Capable of handling varying-length input sequences; however, consistent sequence length stabilizes predictions.

(iii) **Transformer-based models**:

- Utilizes attention mechanisms allowing the model to focus on different parts of the input sequence, irrespective of position.
- Can handle any sequence length up to a maximum set by system memory constraints during training and inference.
- Fixed window size, often referred to as sequence length, should match between training and inference for alignment of positional embeddings.