

Deep Learning and Industrial Applications Homework 3

Name：簡榮信 / Student ID：113034611

(一) 不同視窗大小與步長組合的實驗結果如下表一所示：

表一、window size & step size 不同參數組合下 MSE 結果

編號	Epoch	Window Size	Step Size	Standard MSE	Real MSE
1 (原始)	100	10	15	1.428	1205.644
2	100	10	10	1.468	1310.315
3	100	10	5	1.173	987.373
4	100	5	10	1.462	1200.839
5	100	5	5	1.168	1019.953
6 (最優)	100	5	2	1.096	979.756

實驗分析：

- 共設計六組組合進行比較，固定訓練 100 個 epoch，記錄 Standard MSE 與 Real-scale MSE
- 結果顯示，原始組合 (10,15) 的 MSE 較高，顯示較難捕捉短期變動
- 隨著 window 與 step 降低，模型能擷取更細緻的時間序列變化，預測表現逐步提升
- **最佳組合 (window = 5, step = 2)** 在兩項 MSE 指標皆為最低，代表在此題目中重疊率高的 small step 能提供更豐富的訓練樣本，強化模型對**短期趨勢**的感知能力

🌈 **結論：** 實驗結果顯示，適當縮小 window size 並提高資料重疊率 (減少 step size)，在此題目中可較有效的強化此模型對短期價格變化的感知能力，進而降低預測誤差。

(二)

(i) 加入「Volume」特徵的影響：

A：

🌈 **實驗觀察：**

```
Standardized Test MSE: 1.101243
Real-scale MSE: 966.5352
<ipython-input-81-98e0d4d64d7b>:27: UserWarning:
  mse = F.mse_loss(pred_value, actual_value)
```

- 加入特徵 Volume 後，MSE 由 979 下降至 966，約 **1.3%**

(window size = 5, step size = 2)

🌈 影響分析：

- Volume 提供了額外的補充訊息，幫助模型分辨亦真亦假的趨勢
- 在價格波動劇烈時期會較有效果

🌈 結論：「Volume」是一個有效的輔助特徵欄位，能夠協助強化模型對市場行為的理解與預測準確度。

(ii) 最佳輸入特徵組合探索：

A：

🌈 實驗組合：

- Close 價單獨使用：MSE 最高
- Close + High + Low + Open：MSE = 次之
- Close + High + Low + Volume (最佳組合)：MSE = 最低

🌈 嘗試過的但效果不佳特徵：

```
■ #df['MA5'] = df['Close'].rolling(5).mean()  
■ #df['Momentum'] = df['Close'] - df['Close'].shift(5)  
■ #df['STD5'] = df['Close'].rolling(5).std()
```

- 增加冗餘訊息，訓練不穩定

🌈 分析：

- 高低價提供波動範圍，Volume 補強市場強度認知，形成相對互補效應

🌈 結論：挑選特徵時應避免冗餘訊息，聚焦在能提供不同維度資訊的特徵下，以提升模型泛化能力與精度效果。

(三) 正規化對模型表現的影響：

A：

🌈 觀察結果：

- 未進行正規化時，模型訓練不穩定、收斂速度慢，且預測誤差較高 (Real MSE 約 **1310.3**)
- 使用 StandardScaler 對特徵與標籤進行標準化後，模型收斂更快，預測表現明顯提升

- 最佳正規化模型的實際誤差下降至 Real MSE **979.8**，顯著優於未正規化版本
- Zhu et al. Transformers without Normalization (2025) 文獻指出，儘管部分新架構可移除正規化層，但對於**序列建模與深層網路**，正規化仍能穩定梯度、提升泛化能力
- ✚ **結論：** 正規化仍為各項深度學習模型訓練不可或缺的一步，有助於加速收斂並提升最終預測表現。

(四)為何視窗大小應小於步長？：

A：

✚ **合理性分析：**

- 由第一題的表格結果分析可知，視窗大小在大於步長時表現較為優異，與 Lab 4 所描述的不盡相同。
- 可知**小視窗可能較能更快速反應「局部變化」**
 - ✓ 股票價格具高度波動性，長視窗容易「平滑掉」突變訊號
 - ✓ 小視窗反而能讓模型更專注於「局部波動」，避免資訊被稀釋
- 也可能 LSTM 已內建記憶能力，即使 window size 小，也能透過隱藏狀態維持資訊流

✚ **參考文獻：**

- Lim et al., “Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting” (2021, Nature Machine Intelligence)
- 此文獻模型設計中使用了「sliding window」但不強制使 window < step
- 採用高重疊 (step=1) 視窗來保證時間依賴性
- 證實「短視窗 + 高重疊」能提升模型**短期**預測能力
- ✚ **結論：** 雖然在 Lab 4 中建議視窗大小應小於步長以減少資料重疊，但本次實驗顯示，**較小視窗 (如 window=5) 反而提升模型對短期趨勢的感知能力**。因此，視窗大小的設計應依資料特性與任務目標動態調整。

(五)適用於時間序列的資料增強方法：

A：

✚ **參考方法：** TS2Vec: Towards Universal Representation of Time Series (Yue et al., 2022)

✚ **技術描述：**Time Series to Vector: Self-Supervised Representation Learning for Time Series Forecasting (TS2Vec) 為一種自監督時間序列表示學習方法，透過對不同比例的 window size 進行多尺度的對比學習，增強模型對時間依賴與變化特徵的學習能力。

✚ **優點：**

- 無需額外標註資料即可進行強化學習
- 可有效擷取長短期資訊，提高對未來序列預測的泛化能力
- 支援（此題需求）資料增強做法（如隨機遮蔽、切片、平移等）

✚ **應用領域：**金融、能源預測、生醫訊號處理等多種時間序列任務

✚ **結論：**TS2Vec 提供比傳統 Time Warping 更深層次的**時間依賴強化機制**，也可適用於現代深度強化學習（DRL）時間序列架構。

(六)不同模型之 window size 處理方法：

A：

(i) 卷積模型 (CNN-based)：

- 固定視窗滑動並進行特徵擷取，適合批次輸入與局部特徵建模

(ii) 遞迴模型 (RNN/LSTM-based)：

- 基於時間順序接收相關資料，隱藏層則會動態保留過往歷史訊息
- 適合即時的連續推論

(iii) Transformer 模型：

- 使用固定長度序列（如 512）輸入，亦可搭配 Sliding Window 或 Chunk-wise Attention 優化計算
- 適合處理時間序列較長的推理與全局特徵的捕捉

模型	視窗處理推論	建議 window size	優勢所在
CNN	固定長度視窗	5~30	局部擷取、快且穩定
LSTM	狀態可持續隱藏	10~30	適合連續序列、擅長短期預測
Transformer	固定長度序列	128~512	放眼全局、長期趨勢捕捉



結論： 不同架構對視窗大小策略需求各不相同，理應根據任務性質動態調整，以達到「1.推論效率」與「2.預測表現」兼顧。