National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

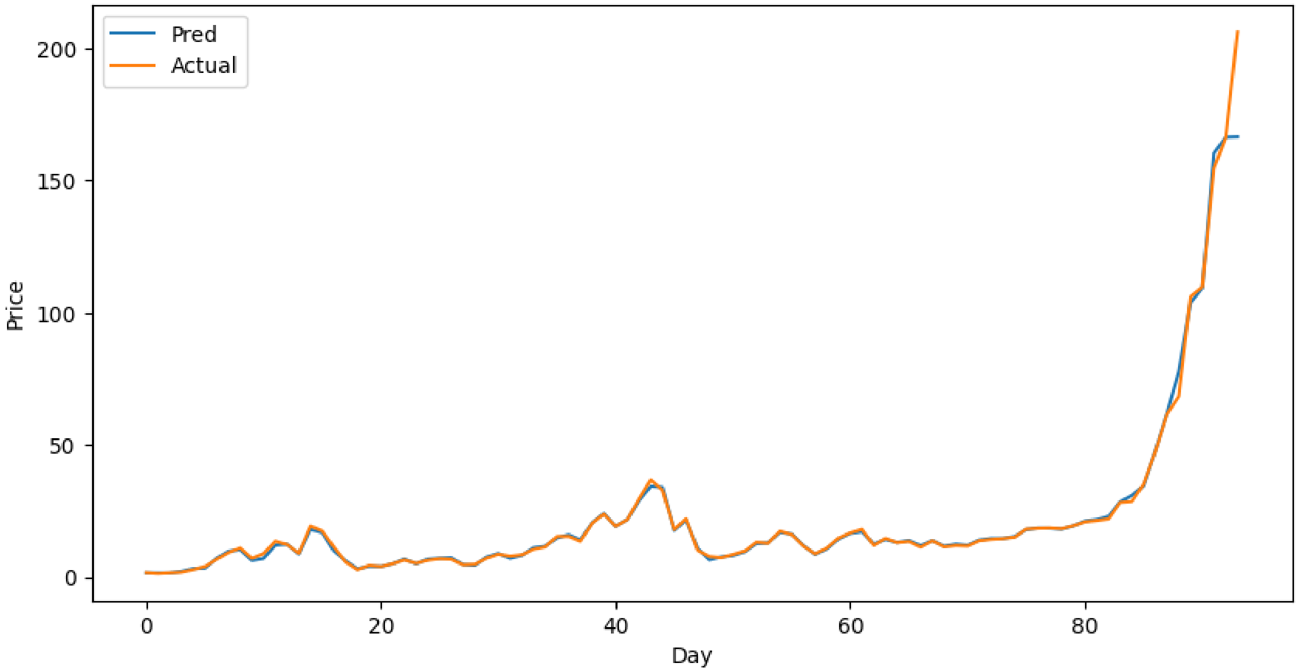
Homework 4

Name:方品心　　　　　　　　Student ID:312704005

Due on 2024/05/02.
Note: DO NOT exceed 3 pages.

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 sizes	steps	Best Val loss
original	10	15	327.0756
1	10	10	77.6164
2	15	10	55.8718
3	15	5	3.0908



改為紅字的組合後，模型預測更準確，如上圖所示。

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.

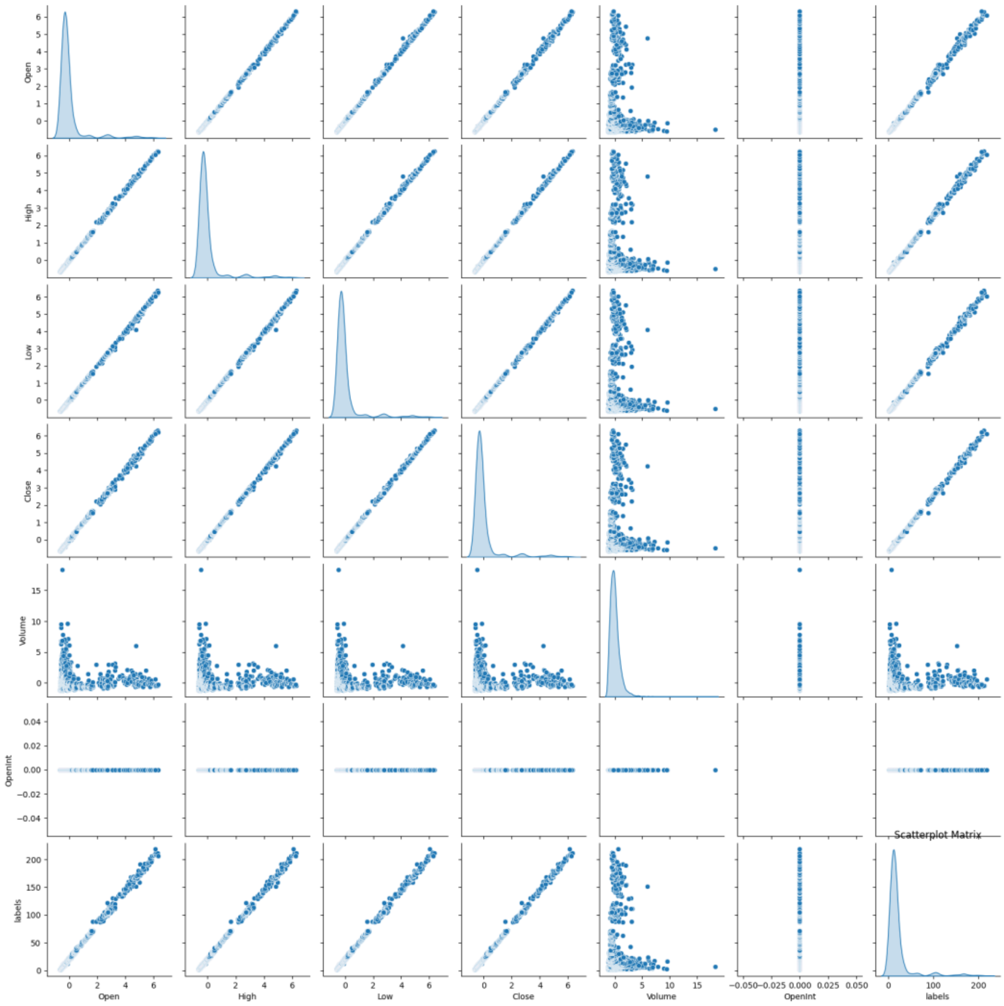
```
Epoch 1/100, Train loss: 1099.0080, Val loss: 550.3309, Best Val loss: 550.3309
Epoch 2/100, Train loss: 968.4555, Val loss: 573.0702, Best Val loss: 550.3309
Epoch 3/100, Train loss: 1007.6660, Val loss: 572.0951, Best Val loss: 550.3309
Epoch 4/100, Train loss: 968.1295, Val loss: 570.2612, Best Val loss: 550.3309
Epoch 5/100, Train loss: 968.7435, Val loss: 562.3740, Best Val loss: 550.3309
Epoch 6/100, Train loss: 969.8711, Val loss: 568.0724, Best Val loss: 550.3309
Epoch 7/100, Train loss: 990.1739, Val loss: 568.0470, Best Val loss: 550.3309
Epoch 8/100, Train loss: 969.7299, Val loss: 567.1605, Best Val loss: 550.3309
Epoch 9/100, Train loss: 968.3184, Val loss: 568.3011, Best Val loss: 550.3309
Epoch 10/100, Train loss: 998.8857, Val loss: 569.4049, Best Val loss: 550.3309
Epoch 11/100, Train loss: 979.2835, Val loss: 568.4953, Best Val loss: 550.3309
Epoch 12/100, Train loss: 968.7663, Val loss: 566.0455, Best Val loss: 550.3309
Epoch 13/100, Train loss: 969.1957, Val loss: 569.1200, Best Val loss: 550.3309
Epoch 14/100, Train loss: 1070.4873, Val loss: 563.8540, Best Val loss: 550.3309
Epoch 15/100, Train loss: 970.3113, Val loss: 572.7393, Best Val loss: 550.3309
Epoch 16/100, Train loss: 970.5774, Val loss: 565.2010, Best Val loss: 550.3309
Epoch 17/100, Train loss: 973.2146, Val loss: 568.1963, Best Val loss: 550.3309
Epoch 18/100, Train loss: 1050.7803, Val loss: 565.3854, Best Val loss: 550.3309
Epoch 19/100, Train loss: 974.7271, Val loss: 572.1256, Best Val loss: 550.3309
Epoch 20/100, Train loss: 1059.3837, Val loss: 569.4986, Best Val loss: 550.3309
Epoch 21/100, Train loss: 971.5649, Val loss: 569.5405, Best Val loss: 550.3309
Epoch 22/100, Train loss: 975.1403, Val loss: 564.4248, Best Val loss: 550.3309
Epoch 23/100, Train loss: 967.5627, Val loss: 564.4737, Best Val loss: 550.3309
Epoch 24/100, Train loss: 971.9038, Val loss: 572.5534, Best Val loss: 550.3309
Epoch 25/100, Train loss: 1058.3733, Val loss: 563.9942, Best Val loss: 550.3309
...
Epoch 97/100, Train loss: 1095.4885, Val loss: 570.0173, Best Val loss: 550.3309
Epoch 98/100, Train loss: 992.9075, Val loss: 570.0141, Best Val loss: 550.3309
Epoch 99/100, Train loss: 965.9955, Val loss: 570.0082, Best Val loss: 550.3309
Epoch 100/100, Train loss: 968.6602, Val loss: 570.0089, Best Val loss: 550.3309
```

加入'Volume' 後模型train不起來，推測是因為Volume裡面的值太大了，與其他特徵的尺度不一樣，會加大影響模型的分析結果。

(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.

我選擇了'Open', 'High', 'Low', 'Close'這四個特徵，跑出的MSE為最低的(如下圖)，並且我有畫出各個特徵之間的散佈圖，這邊的labels是我們想預測的目標(n+1的Hight)，從圖中可以看出它和'Open', 'High', 'Low', 'Close'都高度相關。

```
Epoch 100/100, Train loss: 8.2470, Val loss: 1.7603, Best Val loss: 1.7299
```



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.)

   我認為有 normalization 會提升它的 performance，因為它會統一特徵的維度，讓模型學習的時候比較容易找到最小的loss。那此次作業由於我最後的特徵選擇是'Open', 'High', 'Low', 'Close'四個特徵，他們本身的資料維度就沒有太大的差距，因此有沒有標準化的成果差不多。但如果再加上' Volume '，有沒有標準化就會有明顯的不同。

   ```
   Epoch 100/100, Train loss: 13.3780, Val loss: 7.9830, Best Val loss: 6.4302
   ```

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 window size 必須要大於 step size，首先較大的 window size 可以確保更多的數據點一起被考慮，使模型能夠捕捉更全面的模式和依賴關係，其次，小的 step size 才有較多的訓練資料，這些都增強了模型進行準確預測的能力。

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

   一種對時間序列資料進行增強的方法是時間扭曲(Time Warping)，它改變時間軸的速度，拉伸或壓縮資料。這個技術保留了資料中的時間相依性，同時生成新的樣本，特別適用於時間序列分類或預測等任務。

   References: Wen, Q., Sun, L., Song, X., Gao, J., Wang, X., & Xu, H. (2020). Time Series Data Augmentation for Deep Learning: A Survey. International Joint Conference on Artificial Intelligence.

6. Discuss how to handle window size during inference in different model architectures (approximately 150 words):

   (i) (5 points) Convolution-based models : 主要由 kernel size 控制(和 stride)。通過調整這些參數，可以改變模型對輸入數據的感知範圍。除此之外使用 pooling layers 也可以影響 window size，通過減少空間尺寸來縮小視窗範圍。而 Padding 也是一個重要的參數，它可以用來保持輸入和輸出的尺寸一致，從而影響視窗的實際大小。

   (ii) (5 points) Recurrent-based models : 通常由 sequence length 和 batch processing size 來控制的。對於固定長度的序列，可以將其劃分為固定大小的塊，或者使用 padding 來將序列調整到相同的長度。對於可變長度的序列，可以使用動態填充或 bucketing 技術來處理。

   (iii) (5 points) Transformer-based models : 在 Transformer 模型中，window size 由序列長度參數來控制，這個參數定義了輸入序列中的最大的token數 。對於較長的序列，可以使用分塊或分層方法來處理。Transformer 中的 Self-attention 機制能夠捕捉全局依賴關係，從而在處理大範圍上下文時表現良好。