

National Tsing Hua University
11220IEEM 513600
Deep Learning and Industrial Applications
Homework 4

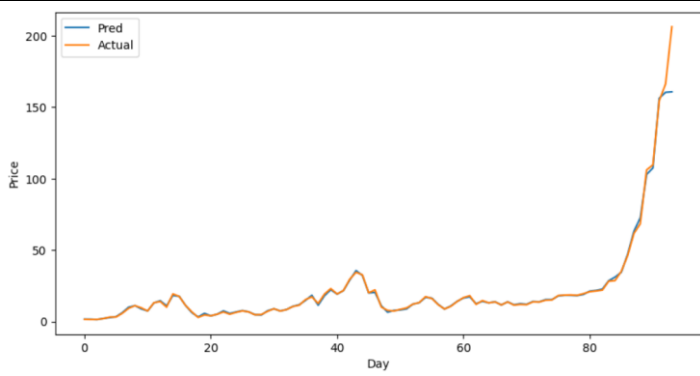
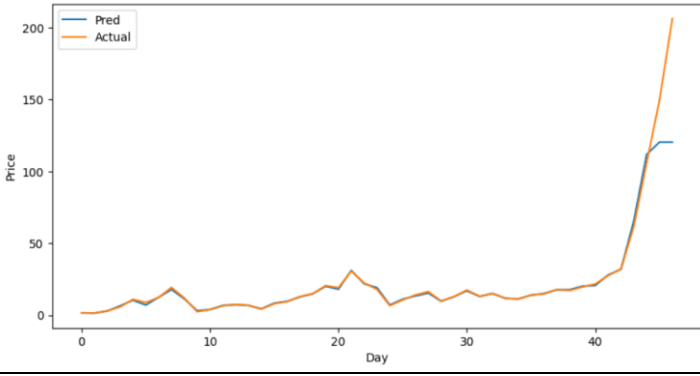
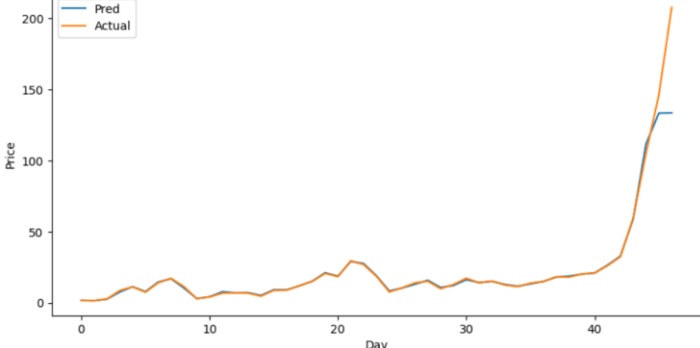
Name: 楊佳蓁

Student ID: 109034048

Due on 2024/05/02.

Note: DO NOT exceed 3 pages.

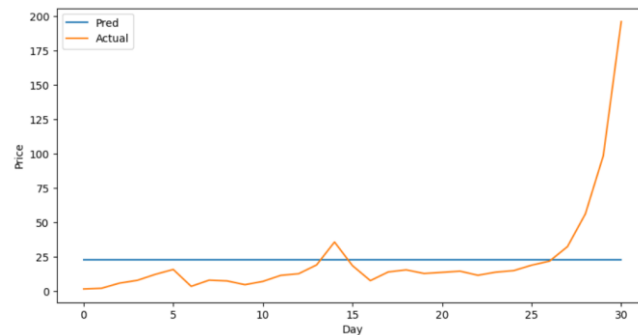
1. (15 points) Experiment with different window sizes and steps. Train the model using **3** different combinations of window size and step. Evaluate the Mean Squared Error (MSE) for each configuration. Report the MSEs using a table and analyze the results. (Approximately 100 words.)

window size / step	Validation Loss	Test Loss	Test Set Prediction
10 / 5	19.05	24.17	
10 / 10	64.09	276.17	
5 / 10	41.18	190.77	

當window size為10且step為5的時候model表現最好，特別是股價飆漲那段。我認為可能是這樣的window size/step組合利用類似移動平均的方式做預測，因此當股價劇烈震盪時能更準確的預測走向。

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.



加入Volume這項feature後，model的表現變得非常糟糕，如上圖是model在test set的表現，股價預測結果呈一直線，且只要feature的組合中有包含Volume，結果都會是這樣。

- (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.

我嘗試的feature組合包含「Open、High、Low、Close」、「Open、Close」、「High、Low」，其中feature組合為「Open、High、Low、Close」時model表現最好也較穩定。由於K棒的組成元素就是開盤價、收盤價、最高價及最低價，常用來表示當天的股價表現，因此我認為用這四項feature預測股價得到較佳結果很合理。

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.)

如果我的feature是選擇「Open、High、Low、Close」，應該不太需要做normalization，因為他們的scale差異不會大的，且model在每個epoch的loss也沒有異常的劇烈震盪。不過我認為如果將「Volume」也加入feature內，就需要normalization，因為成交量與其他feature的scale相差甚遠，normalization可以幫助model更快速學習。

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.)

依照第一題的結果，我認為window size應該要比step小才能讓model表現較佳，因為這樣能夠讓model以類似移動平均的方式預測股價，符合現實中常會用移動平均線(五日線、十日線)觀察股價趨勢的作法。

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

Window warp(Dynamic time warping): 隨機選擇時間範圍並且down-sampling或up-sampling該序列資料，同時保持其他時間範圍不變。這個方法會改變原始時間序列的總長度，所以可搭配window cropping進行，讓資料切成同樣長度的slice。

參考：<https://www.youtube.com/watch?v=G1-AMm7MbIY>

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

(i) (5 points) Convolution-based models

CNN中的window size與kernel的大小相同，根據window size與stride分割input data，接著將每個window的資料進行convolution operation以萃取特徵。

(ii) (5 points) Recurrent-based models

RNN根據window size與step分割input data(序列資料)，再依序利用每段分割資料(大小與window size相同)預測結果。

(iii) (5 points) Transformer-based models

Transformer model的window size代表在attention mechanism中考慮周圍資訊的範圍大小，允許model在處理每個位置的資訊時，同時關注該位置周圍一定範圍內的其他資訊，較大的window size可以捕捉更長的依賴關係，但也增加了運算的複雜度，可透過如local attention mechanism或chunking等方法處理長序列，同時減少計算和記憶體負擔，進而影響window size的選擇。

參考：<https://towardsdatascience.com/de-coded-understanding-context-windows-for-transformer-models-cd1baca6427e>