系所 : 工工所　學號 : 113034511　姓名 : 孫秉新

1.

window sizes	step size	best val loss
10	15	70.0116
5	1	1.5246
10	5	4.1130
30	15	106.6307

一開始的WZ (window size)為10，SZ (step size)為15，得到的結果並不理想，除了預測結果並不好以外，損失函數的下降曲線也不夠平穩。因此，我實驗小、中、大三種WZ和SZ，結果上來看小的表現最好，損失函數的曲線很漂亮，預測結果也相當貼近真實數據，我認為WZ越小可以越反應最新的變動，SZ設小可以有更多的訓練資料，缺點是訓練時間比較久。同時可以觀察到WZ越大，預測結果通常只反映趨勢，所以val loss較大。而中型的之所以表現得比原始的好，是因為SZ較小，所以有較多的訓練資料。

2.

(1)
我使用WZ=5、SZ=1，並加入Volume這個特徵來訓練，得到的結果非常不好，預測的結果如圖1，若沒有加入Volume，預測結果如圖2。所以可以知道Volume這個特徵對預測股價完全沒有幫助，甚至會導致模型無法正常預測。我將各項特徵與隔天股價的相關係數列出來，如圖3，可以知道結果是合理的，因為Volume與股價的像關係數極低。

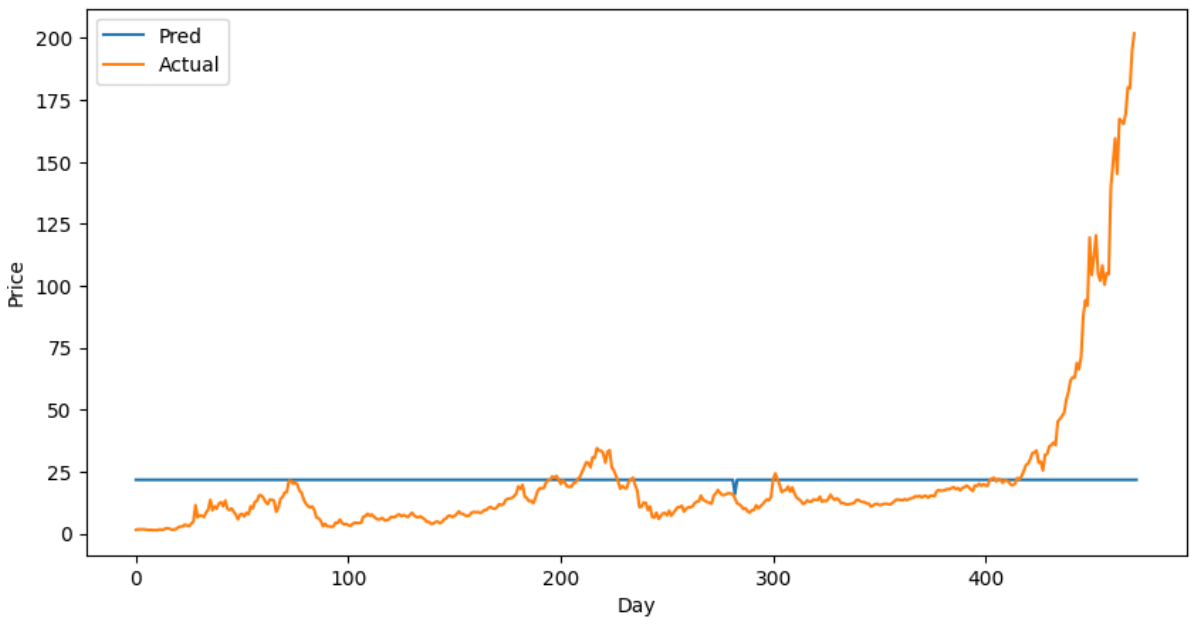

圖1

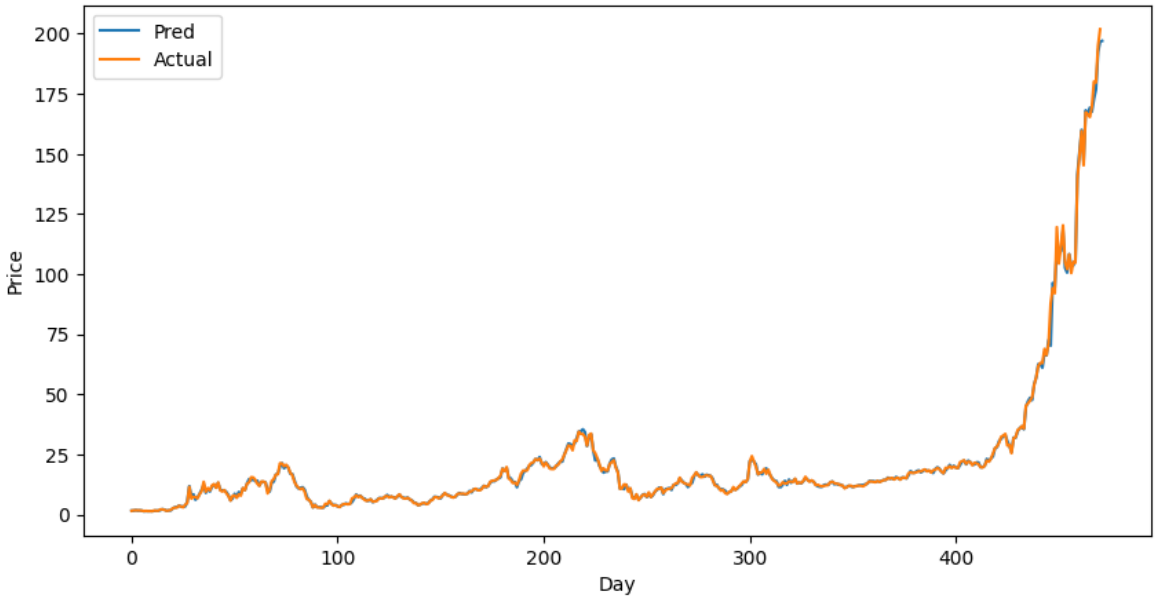

圖2

Open	0.999301
High	0.999553
Low	0.999487
Close	0.999677
Volume	-0.008394
OpenInt	NaN
Next_High	1.000000

圖3

(2)

Open	High	Low	Close	best val loss
有	有	有	無	1.3097
無	有	有	無	1.4240
無	有	有	有	1.8796
無	有	無	無	2.4190
無	無	有	無	1.7777

如表格，我嘗試了以上這些組合，並且val loss最好的是使用Open、High、Low三個特徵，令我比較驚訝的是只使用Low的loss比只使用High的還要低。不過可以發現這幾種組合的loss其

實沒有差異到很大，實際上做的時候我應該還是會四種特徵都採用，因為他們都與股價高度相關。

3.

使用normalize後，我的best val loss是1.5705，比原先的1.5246差了一點，不過並沒有差很多，我認為這可能跟課堂上所學到的normalize有時可能會提供類似正規化的副效果有關，因為正規化相當於對模型添加一些噪音，所以會使得loss稍微上升，但我認為是值得的，因為觀察損失函數下降的曲線，可以發現加入normalize後，模型收斂的速度有變快，也就是可以讓模型更有效率的學習，可以降低訓練所需要的時間。

4.

在部分實驗設計中，將視窗大小設為小於步長有助於增加資料多樣性與減少重複樣本，提高訓練效率。不過根據前面的實驗，我發現在預測股價時，通常視窗大小通常應大於步長，這樣得到的結果都是比較好的。我認為可能是因為在時間序列的資料上若視窗大小小於步長，無法保留時序連貫性。因此，我認為此設定不一定合適，應視預測任務調整。

5.

有一種叫做時間扭曲的時間序列資料增強方法。透過在時間軸上隨機拉伸或壓縮某些區段，使模型學會對時間尺度的變異具有更強的泛化能力。例如，在預測股價時，某些短期波動可被扭曲以模擬市場震盪。這種方式可有效提升模型對於資料變異的容忍度與穩健性。

參考資料: Um, T.T. et al. (2017). Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. ICMI 2017.

6.

(1)

卷積模型通常是透過滑動視窗來處理資料，輸入長度可以大於等於訓練時的視窗大小，因為卷積層本來就可以接收不同長度的資料，只是輸出長度會跟著變。因此，只要保證輸入維度正確，就可以直接丟整段資料進去，不用特別手動切視窗。

(2)

RNN、LSTM、GRU 這種 recurrent models，天生就是一個時間步一個時間步慢慢讀進來的，所以推論時可以處理跟訓練時不同長度的序列。

(3)

Transformer 模型對長度非常敏感。因為它要計算所有 token 之間的 attention，推論時如果太長，可能會出問題或爆記憶體。所以通常做法是：推論時也會限制最大視窗大小（比如訓練時用 512 tokens，推論也限制512），或者用 sliding window 或 chunking 技巧把長序列分段處理。