National Tsing Hua University

11220IEEM 513600

Deep Learning and Industrial Applications

Homework 4

Name:姚國慶　　　　　　　　Student ID:112003802

Due on 2024/05/02.

Note: DO NOT exceed 3 pages.

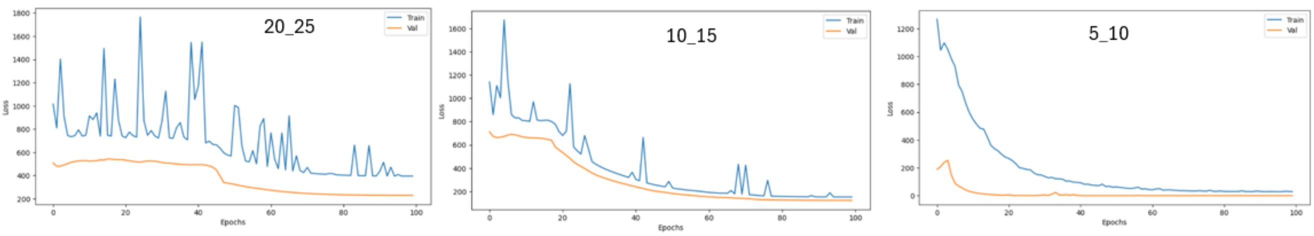
1. (15 points) Experiment with different window sizes and steps. Train the model using **3** different combinations of window size and step. Evaluate the Mean Squared Error (MSE) for each configuration. Report the MSEs using a table and analyze the results. (Approximately 100 words.)

Ans.

三種組合之模型表現MSE如下表。(每種組合MSE為3次平均)

combi-nations	window size	step	batchs	avg val_MSE
1	20	25	189	301.44
2	10	15	315	88.65
3	5	10	473	0.46

各組合之train與validation之loss如下圖

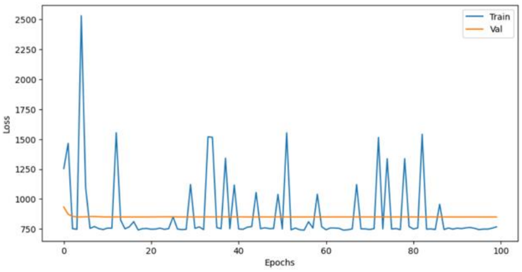


以第3組組合window size 5,step10有最佳的模型表現validation MSE=0.46，比較第1,2組發現當window size愈大MSE表現變差，window size大代表抽樣數變少，訓練資料少，模型偏移(bias)也變得較大。

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.
   (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.

Ans.

(i) 當增加’Volume’作為第5項特徵跑模型時，MSE急遽變差到1125.8559，模型loss無法收斂(如下圖)，表示’Volume’不是預測項’high’的相關性特徵，實務上’Volume’代表是這檔股票的流動性，但無法識別是賣出量或買入量。



(ii) 最好的input features組合是open, high, low, close。設計以下三種(A)、(B)、(C)組合，抽樣數皆設定window size 10,step15

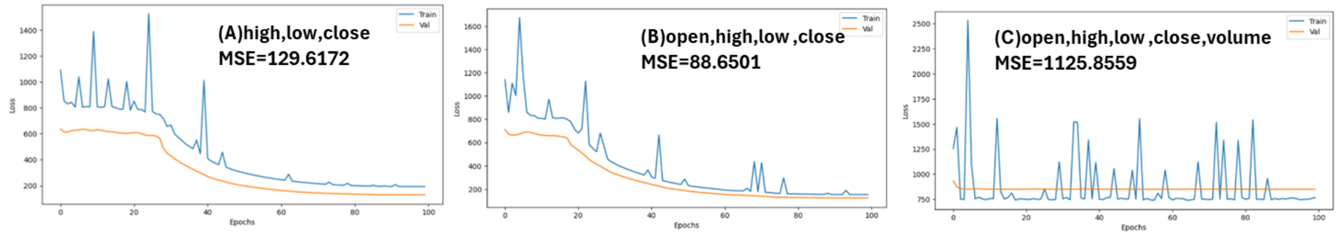
(A) high, low, close

(B) open, high, low, close

(C) open, high, low, close, volume

試驗結果如下圖，(B)的結果最好，沒有overfitting。 (A)也沒有overfitting現象，但Loss較大，(C)則有underfitting，bias大的現象。

對股價預測的特徵重要性依序為close,high,low,open，volume的相關性很低。



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.)

Ans.

實測Lab4,Normalize inputs的效果非常好(採用Min-Max scaling)！ loss幾乎是0 (參下左圖)，實測price時也很準 (參下右圖)。(註：code未做demormlization，所以price的scale仍在0-1之間)

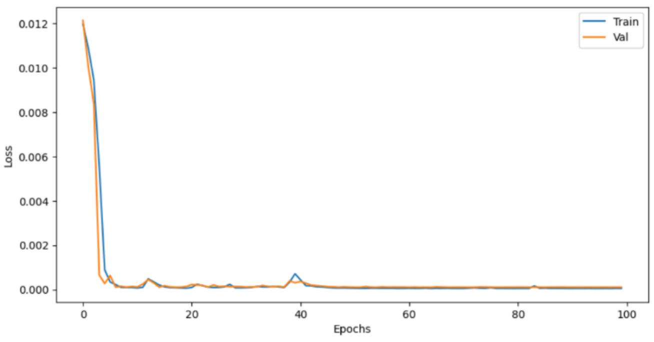
```
[12]  # Select the columns containing the inputs
      input_columns = ['Open', 'High', 'Low', 'Close', 'Volume']
```

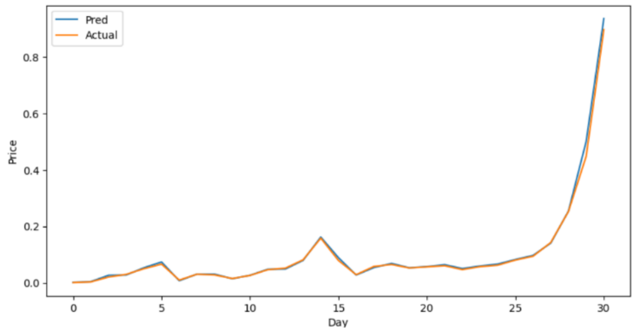
```
# Normalize the inputs using Min-Max scaling
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
df[input_columns] = scaler.fit_transform(df[input_columns])

# Display the normalized DataFrame
print(df)
```

	Date	Open	High	Low	Close	Volume	OpenInt
0	1999-01-22	0.001534	0.002262	0.000970	0.001208	0.073288	0
1	1999-01-25	0.001621	0.001751	0.001367	0.001942	0.013800	0
2	1999-01-26	0.001884	0.001921	0.001413	0.001338	0.009296	0
3	1999-01-27	0.001226	0.001280	0.001103	0.001338	0.006660	0
4	1999-01-28	0.001183	0.001109	0.001413	0.001295	0.006169	0
...	...	...	...	...	...	...	...
4728	2017-11-06	0.972235	0.960019	0.976568	0.969704	0.038979	0
4729	2017-11-07	0.988054	0.973453	0.992519	0.980733	0.042744	0
4730	2017-11-08	0.994192	0.969313	0.979135	0.967516	0.052205	0
4731	2017-11-09	0.963122	0.943226	0.946482	0.949646	0.095708	0
4732	2017-11-10	1.000000	1.000000	1.000000	1.000000	0.125371	0

[4733 rows x 7 columns]





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.)

Ans.

window size	step	batchs	avg_val_MSE
5	5	946	3.96
	10	473	0.46
10	10	473	101.26
	15	315	88.65
20	20	236	702.86
	25	189	301.44

從以上模型測試可以看出，當window size < step size時模型表現較好。因為Lab4 dataset是18年長期的資料，所以預測主要要抓到資料的趨勢，設計window size < step size之參數，時序資料不會重疊，有助於抓到資料內含的趨勢。

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

Ans.

Time warping. Time warping alters the time scale of the series without changing its content, introducing variability while preserving temporal relationships.

Dynamic Time Warping (DTW) is commonly used for aligning time series by stretching or compressing them to minimize the distance between corresponding points. This technique is beneficial when dealing with time-series data with varying speeds or temporal distortions. Time warping has been applied in various fields, including speech recognition, signature verification, and time-series analysis.

Reference: Fulcher, B. D., & Jones, N. S. (2014). Highly comparative feature-based time-series classification. Knowledge and Data Engineering, IEEE Transactions on, 26(12), 3026-3037.

6. Discuss how to handle window size during inference in different model architectures (approximately 150 words):
   (i) (5 points) Convolution-based models
   (ii) (5 points) Recurrent-based modelsse
   (iii) (5 points) Transformer-based models

Ans.

(i) Convolution-based models, **the window size is typically determined by the**

kernel size of the convolutional filters. During inference, input data is fed into the model using the same fixed window size as during training.

(ii) recurrent-based models, **the window size is determined by the length of the input sequence**. During inference, the model processes input sequences one timestep at a time, maintaining an internal state that captures temporal dependencies.

(iii) transformer-based models, the handling of window size during inference involves dividing the input sequence into **fixed-length segments** or "windows" during both training and inference. These segments are often referred to as **"tokens"** and are typically determined by the **maximum sequence length** parameter set during model training. During inference, the model generates predictions by attending to all tokens within the window, utilizing self-attention mechanisms to capture long-range dependencies effectively.