

中山大学

时间序列分析课程报告

题目：使用 LSTM 对股票收盘价进行预测

姓 名： 吴梓彬

学 号： 21311594

院 系： 人工智能学院

任课教师： 谷德峰

中山大学人工智能学院

2023 年 11 月 20 日

目录

1. 引言	3
2. 数据分析	3
2.1 原始序列非平稳.....	3
2.2 一阶差分后的收盘价是平稳白噪声序列.....	4
3. 方法	5
3.1 LSTM	5
4. 结果分析	6
5. 总结	7

1. 引言

数据来源为课程提供的股票数据，预测的目标为次日的收盘价，工作内容是使用 LSTM 网络预测股票收盘价。

2. 数据分析

2.1 原始序列非平稳

在 ACF 图中，自相关系数随着滞后的增加而缓慢下降，没有快速的衰减，这表明序列存在长期的相关性，PACF 中的自相关系数随着滞后的增加而快速的下降到置信区间内，这表明序列具有自回归特性，但不是简单的自回归过程。因此，根据 ACF 和 PACF 的分析结果，该股票收盘价序列很可能是非平稳的。

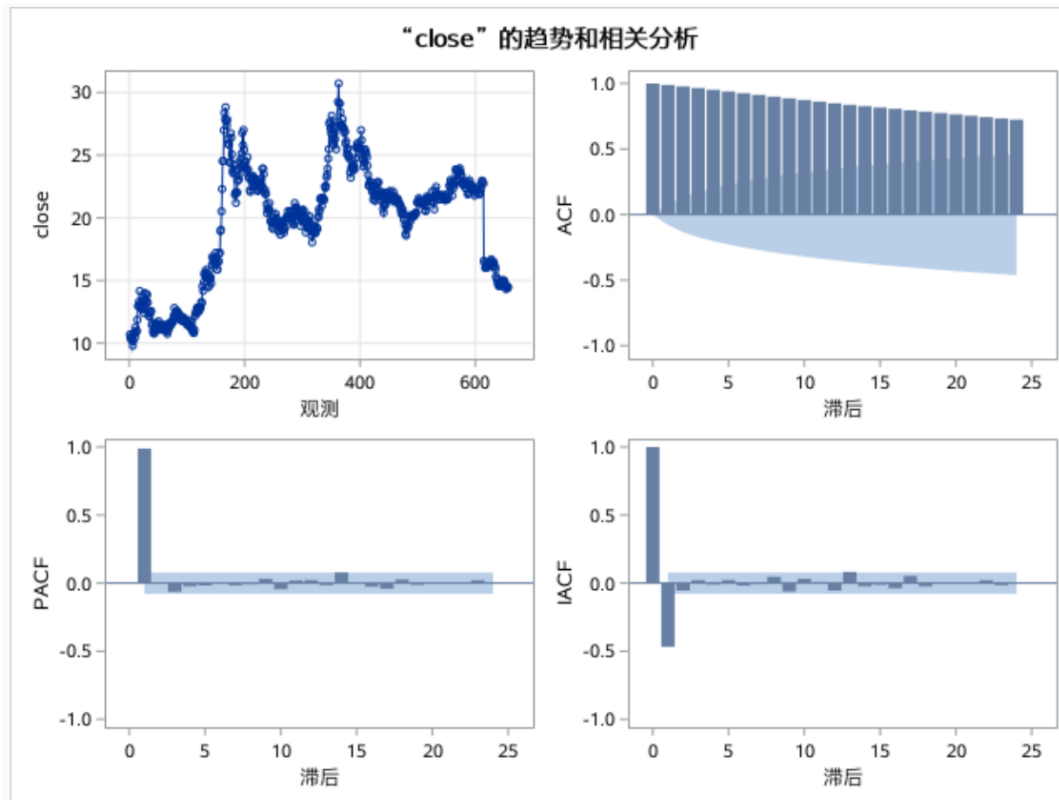


图 2.1.1 收盘价的趋势和相关分析图

因为 ADF 检验中 P 值均大于 0.05，故无法拒绝原假设，即序列非平稳。

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.1712	0.6439	-0.23	0.6053		
	1	-0.1502	0.6487	-0.20	0.6134		
	2	-0.1905	0.6395	-0.24	0.6007		
Single Mean	0	-6.3681	0.3183	-1.98	0.2960	1.99	0.5609
	1	-6.2379	0.3281	-1.98	0.2965	1.99	0.5602
	2	-7.2554	0.2585	-2.11	0.2405	2.26	0.4909
Trend	0	-5.0864	0.8139	-1.38	0.8656	2.21	0.7334
	1	-4.8142	0.8334	-1.34	0.8778	2.27	0.7211
	2	-6.0366	0.7408	-1.54	0.8158	2.41	0.6929

图 2.1.2 收盘价的 ADF 检验

2.2 一阶差分后的收盘价是平稳白噪声序列

白噪声的自相关检查中，P 值均大于 0.05，故无法拒绝原假设，为白噪声序列。因为 ADF 检验中 P 值均小于 0.05，故拒绝原假设，即序列平稳。

白噪声的自相关检查									
至滞后	卡方	自由度	Pr > 卡方	自相关					
6	6.46	6	0.3738	-0.028	0.079	0.031	0.003	-0.022	0.036
12	10.74	12	0.5512	0.022	-0.044	0.035	-0.017	-0.047	0.016
18	20.54	18	0.3033	-0.096	0.032	0.018	0.058	-0.019	-0.014
24	21.12	24	0.6317	-0.018	0.000	-0.004	-0.011	0.008	-0.019

增广 Dickey-Fuller 单位根检验							
类型	滞后	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-672.956	0.0001	-26.29	<.0001		
	1	-574.386	0.0001	-16.92	<.0001		
	2	-517.620	0.0001	-13.58	<.0001		
Single Mean	0	-673.021	0.0001	-26.27	<.0001	345.04	0.0010
	1	-574.566	0.0001	-16.91	<.0001	143.05	0.0010
	2	-517.898	0.0001	-13.58	<.0001	92.15	0.0010
Trend	0	-675.680	0.0001	-26.36	<.0001	347.34	0.0010
	1	-581.479	0.0001	-17.01	<.0001	144.60	0.0010
	2	-529.051	0.0001	-13.67	<.0001	93.43	0.0010

图 2.2.1 收盘价一阶差分的白噪声检验和 ADF 检验

3. 方法

3.1 LSTM

LSTM 网络是一种循环神经网络，它可以利用长短期记忆单元来处理时间序列数据，如股票价格。LSTM 网络可以学习股票价格的历史走势和波动规律，从而对未来的收盘价进行预测。使用 LSTM 网络预测股票收盘价的工作内容大致可以分为以下几个步骤：

1. 数据准备：首先，需要从课程提供的股票数据中提取出所需的特征，提取了开盘价、收盘价、最高价、最低价、成交量、成交总额。然后，需要对数据进行预处理，进行了归一化、划分训练集和测试集、使用 lookback 天预测 1 天。最后，需要将数据转换为 LSTM 网络所需的输入格式，即一个三维的张量，其中第一维表示样本数，第二维表示时间步长为 lookback，第三维表示特征数为 6。
2. 模型构建：使用 Pytorch 深度学习框架来构建 LSTM 网络的模型。该 LSTM 网络的模型由 3 个 LSTM 层和全连接层组成，其中 LSTM 层负责提取时间序列数据的特征，全连接层负责输出预测值。此外，还添加了 Dropout 层，以提高模型的泛化能力和稳定性。
3. 模型训练：接着，需要使用训练集来训练 LSTM 网络的模型。在训练过程中，选择的损失函数为 MESLoss、优化器为 Adam、学习率为 0.001，以使模型能够有效地学习数据的特征，并在验证集上达到最佳的性能。同时，还需要监控模型的训练过程，如损失曲线，以便及时调整模型的参数或结构。

- 模型评估：最后，需要使用测试集来评估 LSTM 网络的模型。在评估过程中，选择的评估指标为均方误差，以衡量模型的预测效果。同时，还使用可视化的折线图，来展示模型的预测值和真实值的对比，以直观地观察模型的优劣。

4. 结果分析

多次训练模型得到的最优结果为，在测试集上的 $MSE=0.39$ ，该结果十分接近一阶滞后的测试数据与原始测试数据之间的均方误差，且由图 4.4 中也可看出预测数据滞后于真实数据。同时也对一阶差分后的收盘价（即 change）进行多次了训练和预测，由图 4.6 可以看出 LSTM 网络对 change 值预测效果很差，因其为白噪声序列。产生滞后问题的根本原因是数据序列中产生了变化趋势（或者说是非线性非平稳序列），而且股票序列中随机变化的趋势频繁出现，且训练数据中找不到相应的规律，那么在这个问题上使用过去的的数据预测未来是无解的，或者说几乎是不可预测的。

```
Train Score: 0.39 MSE
Test Score: 0.39 MSE
```

图 4.1 模型在训练集与测试集上的均方误差

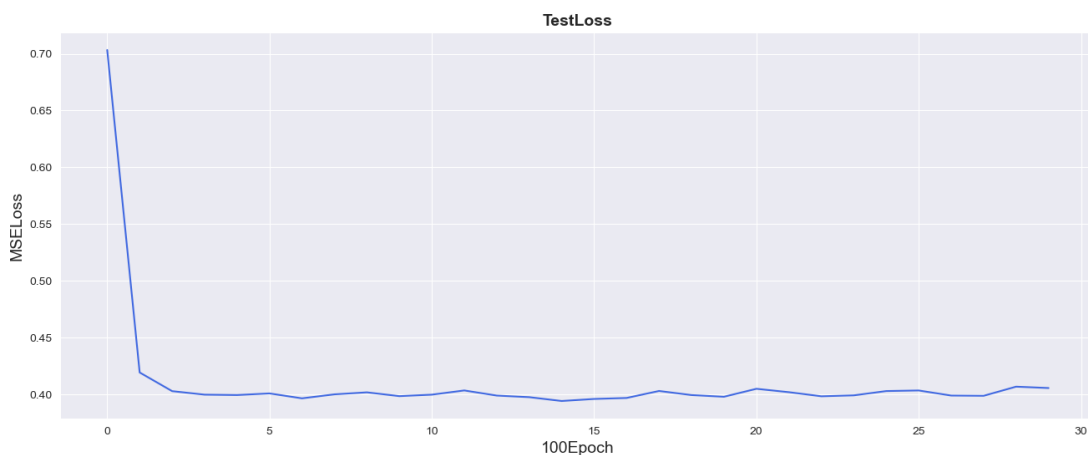


图 4.2 训练过程中模型在测试集上的均方误差曲线

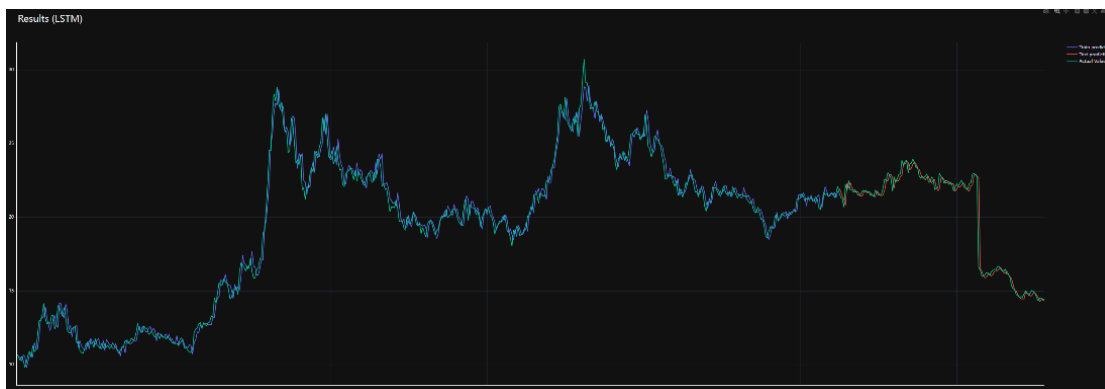


图 4.3 原始数据及预测结果图（所有）

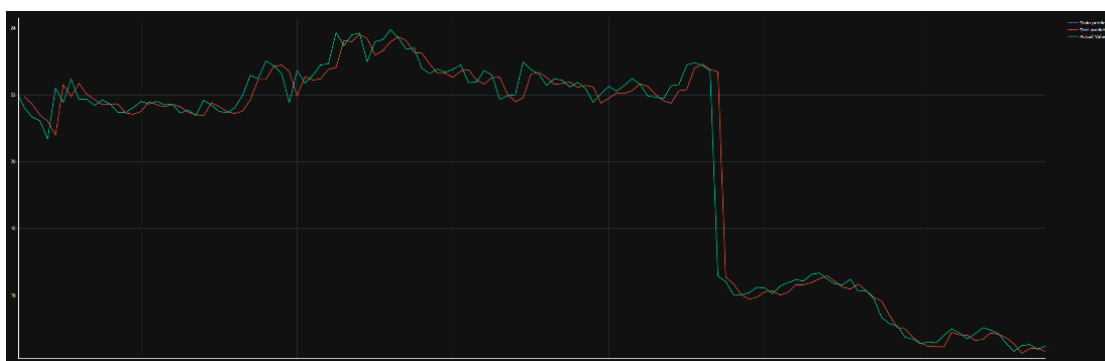


图 4.4 原始数据及预测结果图（测试集）

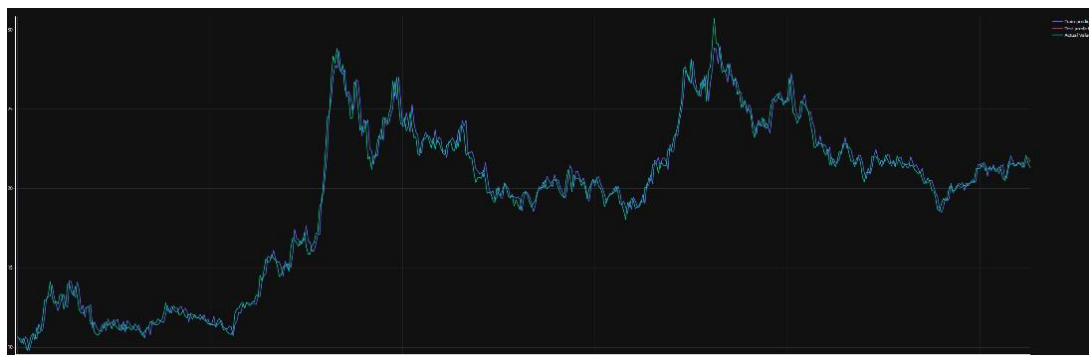


图 4.5 原始数据及预测结果图（训练集）

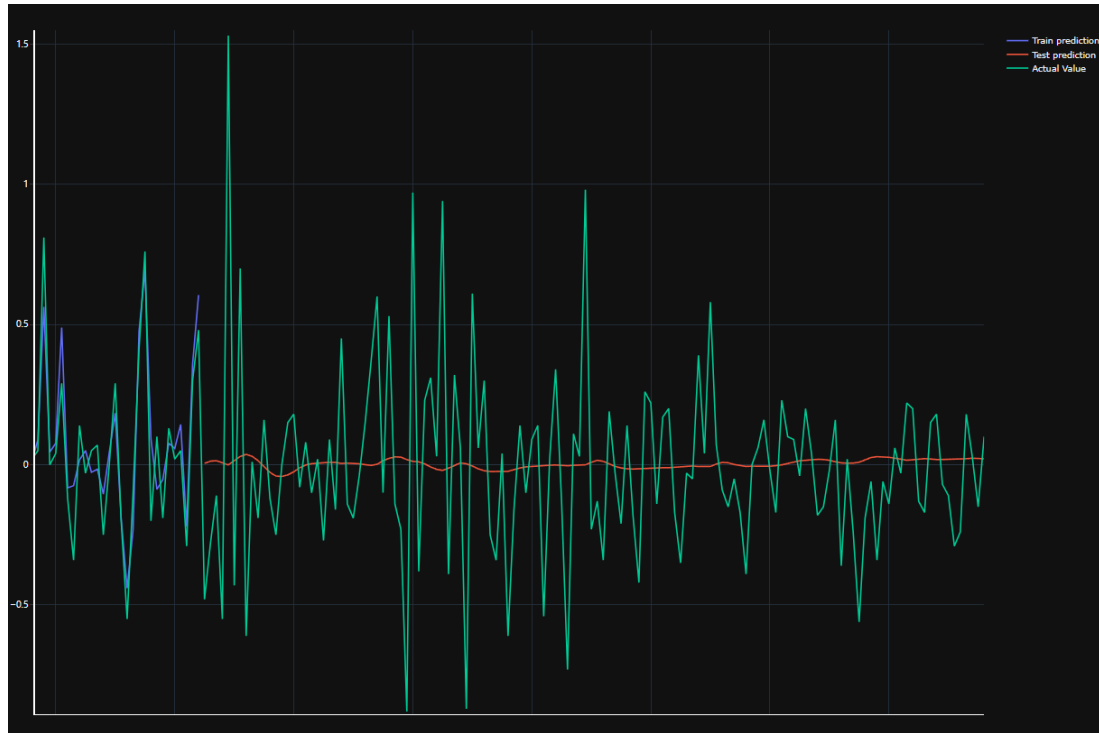


图 4.6 一阶差分的原始数据及预测结果图

5. 总结

使用 LSTM 进行股票预测是一个热门的话题，也是一个有争议的问题。LSTM 是一种深度学习模型，可以处理时间序列数据，捕捉长期的依赖关系。理论上，LSTM 可以从历史的股价数据中学习出隐藏的规律，从而对未来的股价走势做出预测。然而实际上，LSTM 预测股价的效果并不理想，有以下几个原因：

1. 股价受到很多因素的影响，如市场情绪、宏观经济、政策变化、突发事件等，这些因素往往是难以量化的，也是难以预测的。LSTM 只能从历史数据中学习，而不能考虑这些复杂的外部因素。
2. 股价本身具有很强的随机性和非线性性，很难用一个确定的模型来描述。

LSTM 可能会过度拟合历史数据，而忽略了股价的不确定性和变化性。LSTM

的预测结果往往与前一天的实际值非常接近，但是并没有捕捉到股价的波动和趋势。

3. 股价预测是一个竞争激烈的领域，很多投资者和机构都在使用各种方法和策略来获取利润。LSTM 的预测结果可能会被其他人利用，从而影响股价的走势。LSTM 的预测结果也可能会与其他方法的预测结果相互冲突，从而造成混乱和误导。

综上所述，LSTM 可能在某些场景下有一定的作用，但是也需要注意其局限性和风险。LSTM 并不是一个万能的工具，也不能替代人的判断和思考。