

1. 项目基本信息

项目名称	股票市场预测研究
开始日期	2016.5.15
结束日期	2016.6.26
小组成员:	黄姗、靳晓乐、马骁

2. 项目综述

2.1 任务目标

通过处理及分析股票技术关键指标，建立相应的预测模型，实现股票第二天收盘价预测。

2.2 选题意义

股票市场是一个国民经济发展变化的“晴雨表”，股票价格的涨跌也是政治、经济、社会诸多因素的综合反映。股票市场的预测研究是对整个市场的走势和风险进行分析，提前预判组合资产面临的收益和风险。

2.3 主要工作

本次实验通过分析股票技术指标数据，分别训练神经网络拟合模型、神经网络时间序列模型和 ARIMA 模型实现股票收盘价预测。

3. 背景知识

3.1 股票技术指标

股票技术指标有很多，主要的一些技术指标有：

多空指数(BBI)

指数平滑移动平均线(MACD)

相对强弱指数(RSI)

能量潮(OBV)

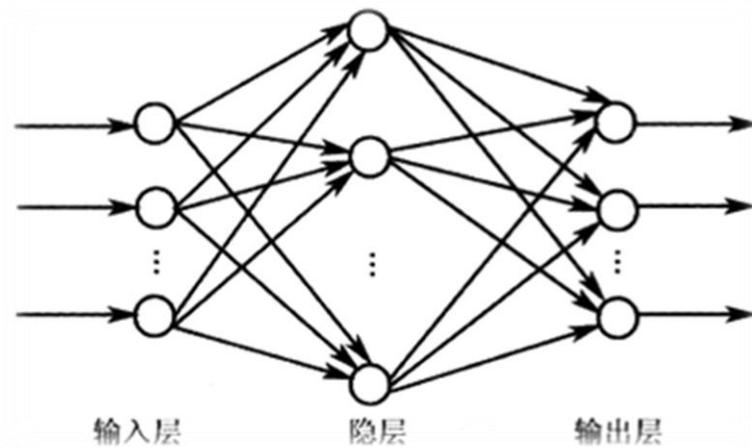
移动平均线(MA)

威廉指数(WMS)

CR 指标(CR)

3.2 人工神经网络

人工神经网络（artificial neural network，缩写 ANN），简称神经网络（neural network，缩写 NN），是一种模仿生物神经网络的结构和功能的数学模型或计算模型。神经网络由大量的人工神经元联结进行计算。结构如下图所示：



3.3 Arima 模型

ARIMA (p, d, q) 自回归积分滑动平均模型 (Autoregressive Integrated Moving Average Model), 又称为差分自回归移动平均模型, AR 是自回归, p 为自回归项; MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。

4. 项目执行情况

4.1 数据处理部分

4.1.1 原始数据来源: 数据来自三个证券交易所

上证 000001: 从 1990 年 12 月 19 日到 2016 年 6 月 01 日

深证 399001: 从 1991 年 04 月 03 日到 2016 年 6 月 01 日

沪深 399300: 从 2002 年 01 月 04 日到 2016 年 6 月 01 日

原始数据结构如下 (前十条):

1	日期	股票代码	名称	收盘价	最高价	最低价	开盘价	前收盘	涨跌额	涨跌幅	成交量	成交金额
2	2016/6/1	'000001	上证指数	2913.508	2929.076	2909.512	2917.154	2916.616	-3.1083	-0.1066	188386421	2.20E+11
3	2016/5/31	'000001	上证指数	2916.616	2917.135	2822.593	2822.593	2822.451	94.1651	3.3363	215260341	2.37E+11
4	2016/5/30	'000001	上证指数	2822.451	2830.97	2794.661	2814.651	2821.046	1.4049	0.0498	106319589	1.16E+11
5	2016/5/27	'000001	上证指数	2821.046	2832.799	2809.799	2817.968	2822.443	-1.3969	-0.0495	109845823	1.24E+11
6	2016/5/26	'000001	上证指数	2822.443	2827.092	2780.763	2813.543	2815.086	7.3566	0.2613	114766797	1.27E+11
7	2016/5/25	'000001	上证指数	2815.086	2843.165	2807.749	2835.029	2821.666	-6.58	-0.2332	103527265	1.18E+11
8	2016/5/24	'000001	上证指数	2821.666	2839.695	2807.188	2839.682	2843.645	-21.9787	-0.7729	111452445	1.21E+11
9	2016/5/23	'000001	上证指数	2843.645	2848.071	2826.256	2826.312	2825.483	18.1619	0.6428	120470455	1.40E+11
10	2016/5/20	'000001	上证指数	2825.483	2825.952	2785.08	2792.888	2806.906	18.577	0.6618	108700899	1.24E+11

原始记录包括日期、股票代码、名称、收盘价、最高价、最低价、开盘价等。过滤后留下收盘价、最高价、最低价、开盘价和成交量。

4.1.2 股票技术关键指标数据的获取：

候选技术指标：

MACD、WMS、RSI、OBV、M30、M10、K 指标、D 指标、DIF

DEA、收盘价

实验确定关键技术指标

4.2 神经网络拟合模型

4.2.1 神经网络配置

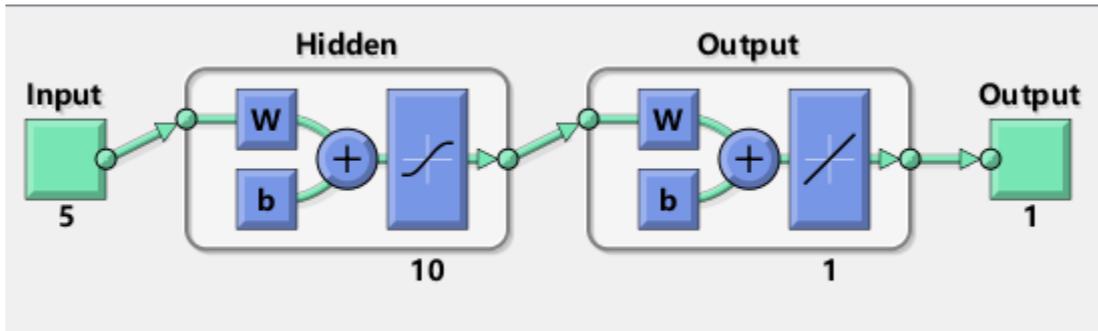
根据选择的技术指标的不同，有不同的输入神经元个数，从 5 个到 11 个不等，输出节点为 1 个，调整隐含层结点进行实验。

激活函数隐含层选用逻辑回归，输出层选择线性回归；

性能评价选用 MSE 以及准确率；

指定训练集为 70%，验证集和测试集均为 15%

由以上配置，得到神经网络如下：



如图所示是输入神经元个数为5,隐含层个数为10时的神经网络。

4.2.2 实验结果

A.

1、输入指标:

MACD、WMS、RSI、OBV、收盘价、M30、M10、K 指标、D 指标

DIF、DEA

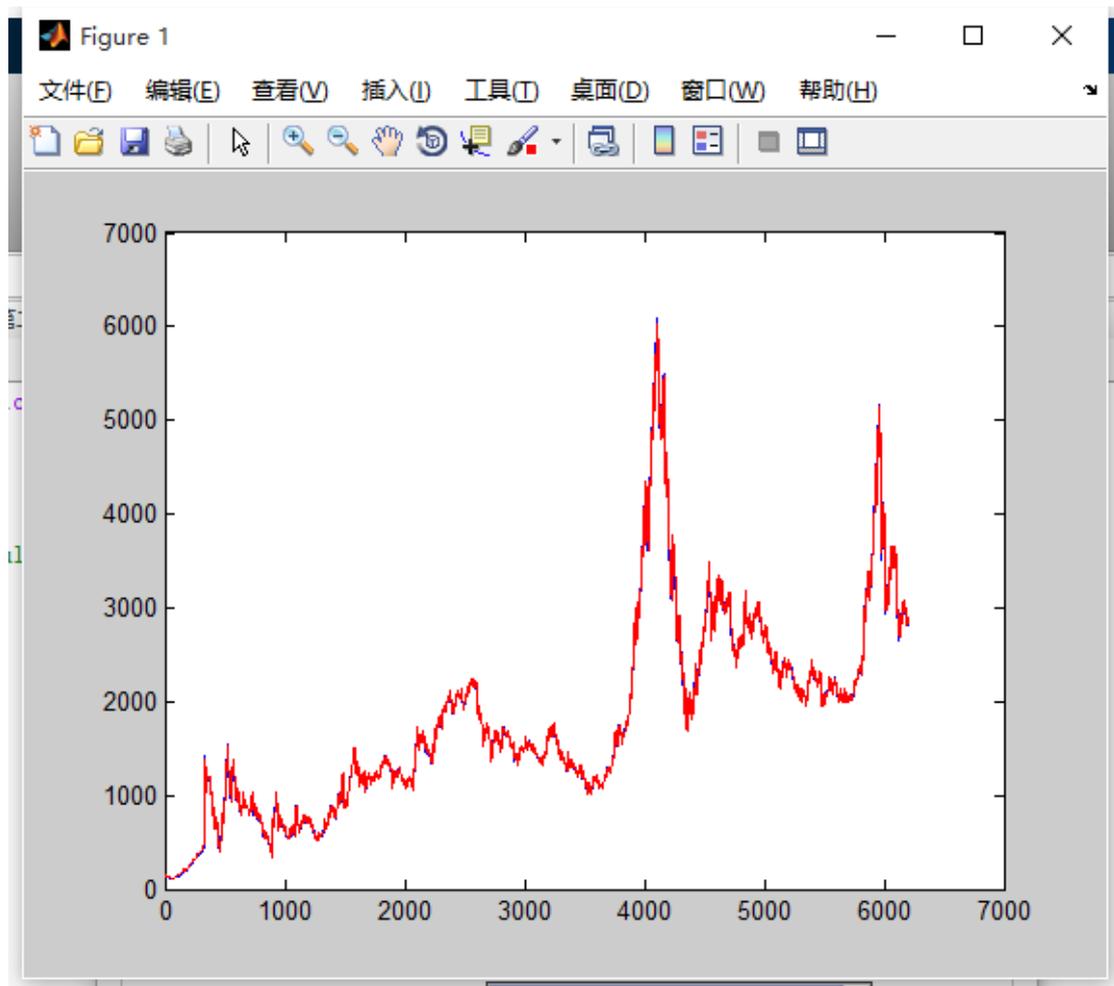
2、输出:

第二天收盘价

3、隐含层个数:

10

4、拟合结果:



0.01 以内的误差准确率: 0.495968

0.02 以内的误差准确率: 0.737258

0.05 以内的误差准确率: 0.930968

涨跌准确率: 0.886485

B.

1、输入指标:

MACD、WMS、RSI、OBV、收盘价、M30、M10、K 指标、D 指标

DIF、DEA

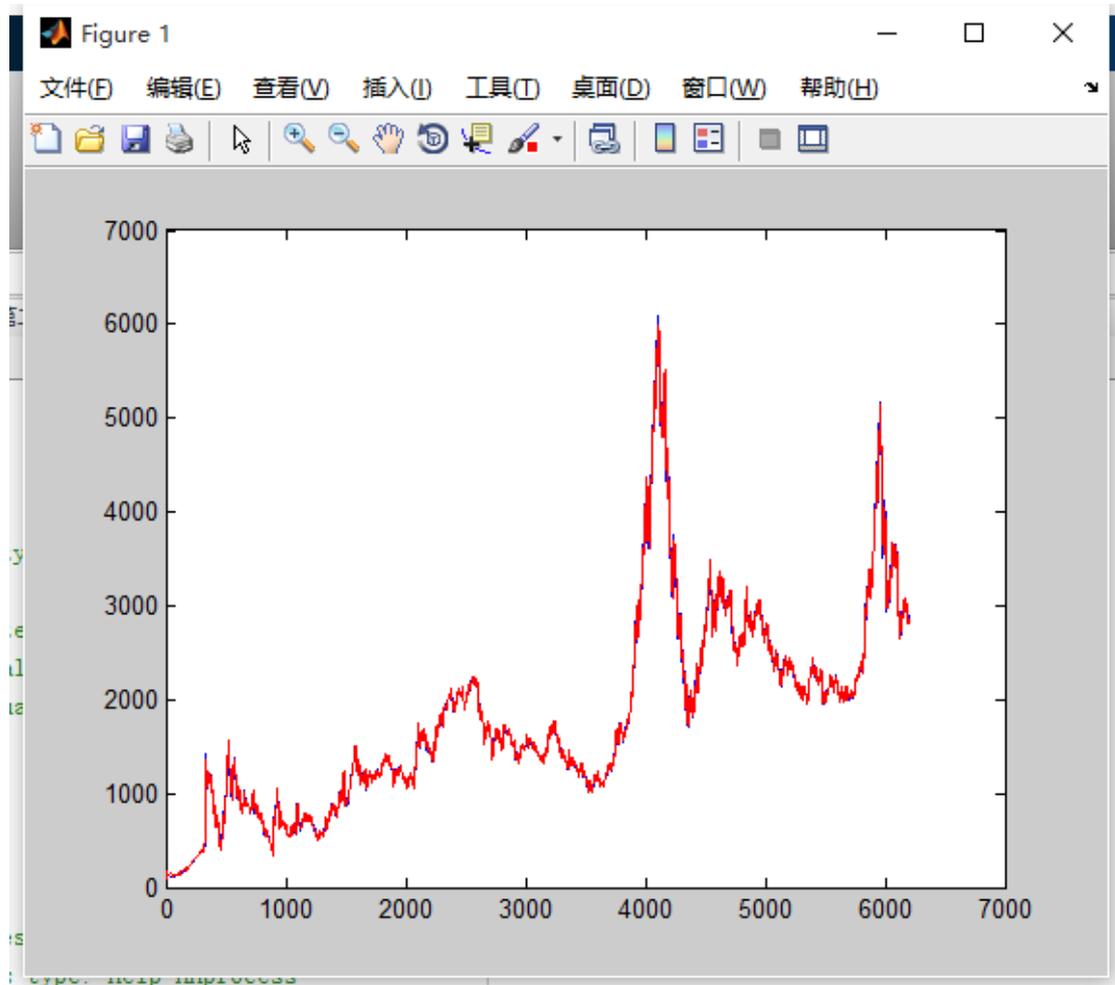
2、输出:

第二天收盘价

3、隐含层个数:

23

4、拟合结果:



0.01 以内的误差准确率: 0.499355

0.02 以内的误差准确率: 0.747903

0.05 以内的误差准确率: 0.948065

涨跌准确率: 0.888422

C.

1、输入指标:

MACD、WMS、RSI、OBV、收盘价、M30、M10、K 指标、D 指标

DIF、DEA

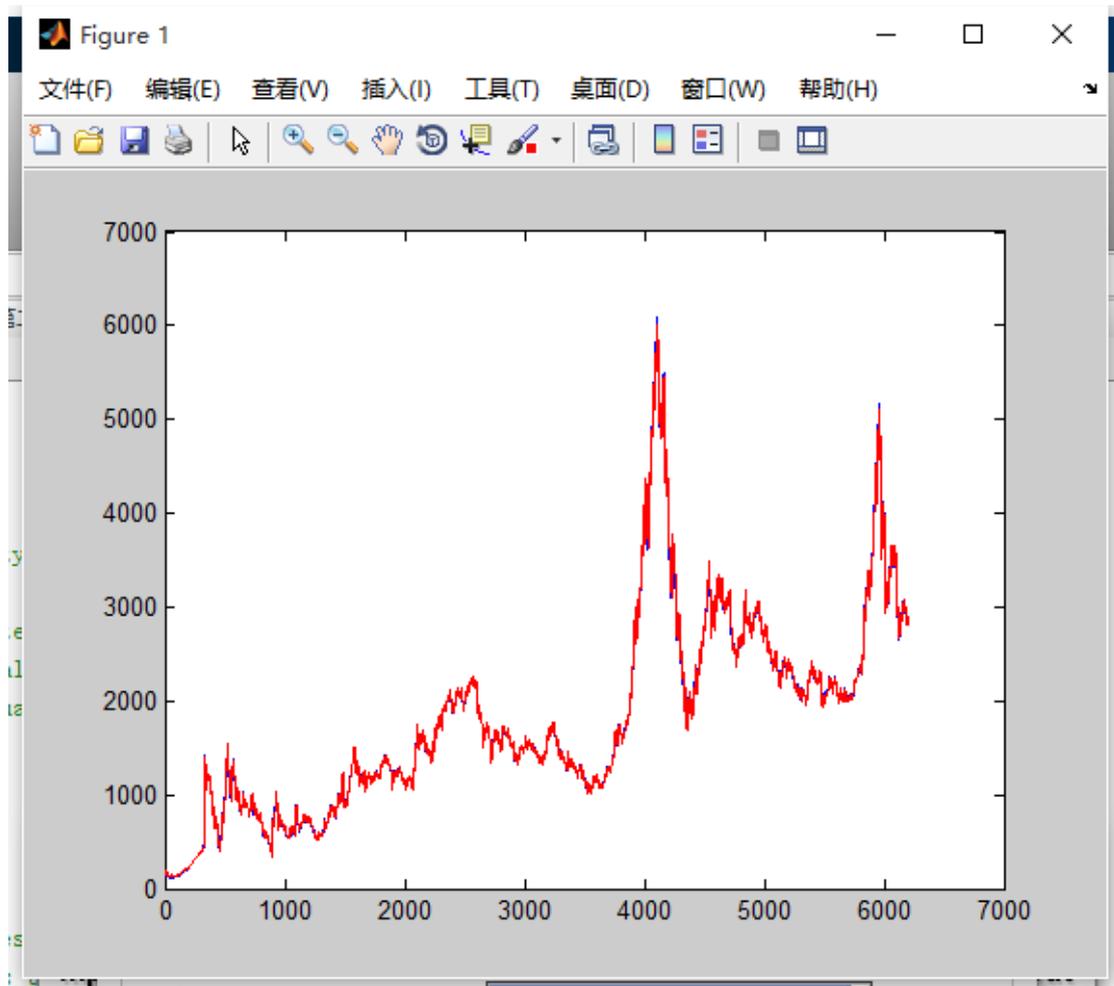
2、输出：

第二天收盘价

3、隐含层个数：

3

4、拟合结果：



0.01 以内的误差准确率： 0.525484

0.02 以内的误差准确率： 0.763226

0.05 以内的误差准确率： 0.938065

涨跌准确率： 0.880995

D.

1、输入指标:

MACD、WMS、RSI、OBV、收盘价、M30、M10、K 指标、D 指标

DIF、DEA

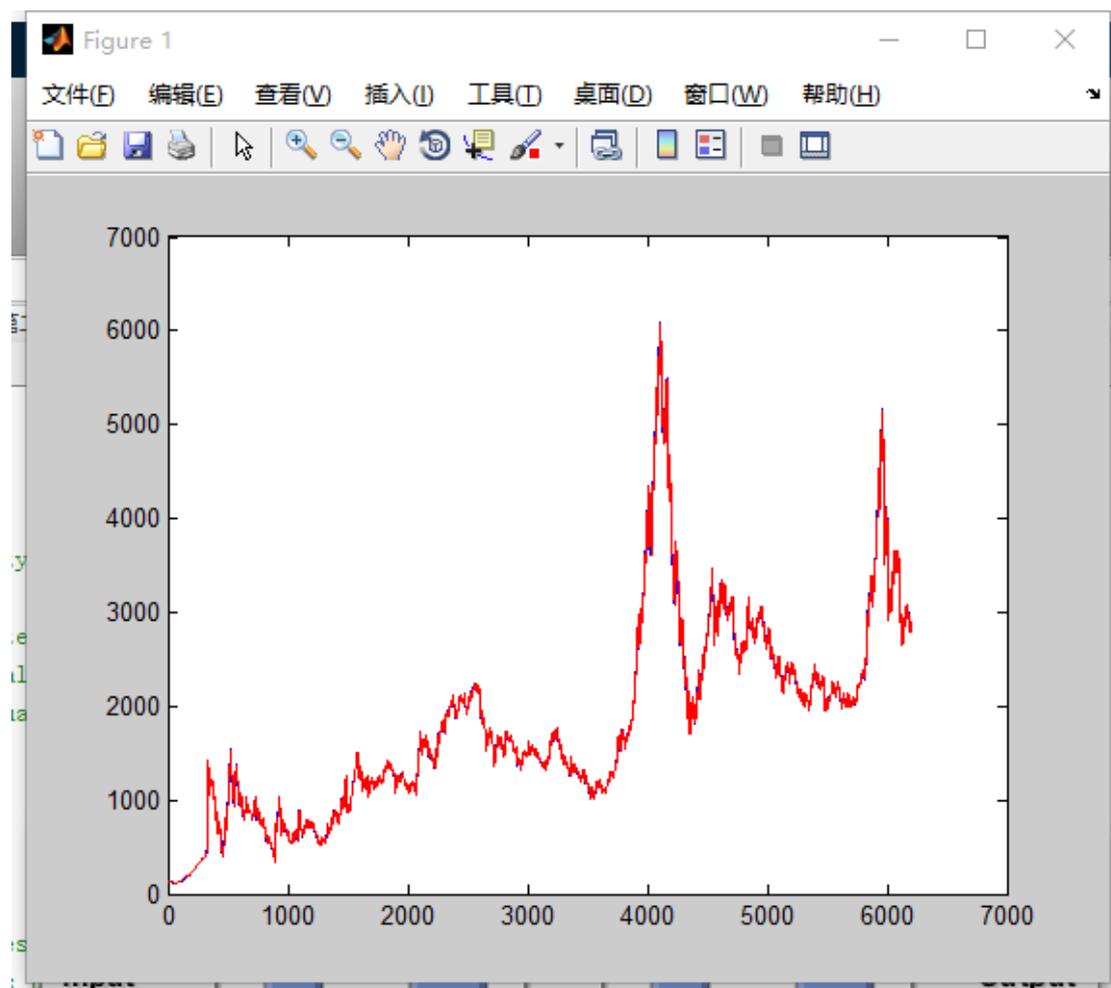
2、输出:

第二天收盘价

3、隐含层个数:

4

4、拟合结果:



0.01 以内的误差准确率: 0.533065

0.02 以内的误差准确率： 0.770806

0.05 以内的误差准确率： 0.958710

涨跌准确率： 0.889714

E.

1、输入指标：

MACD、WMS、RSI、OBV、收盘价、M30、M10、K 指标、D 指标

DIF、DEA

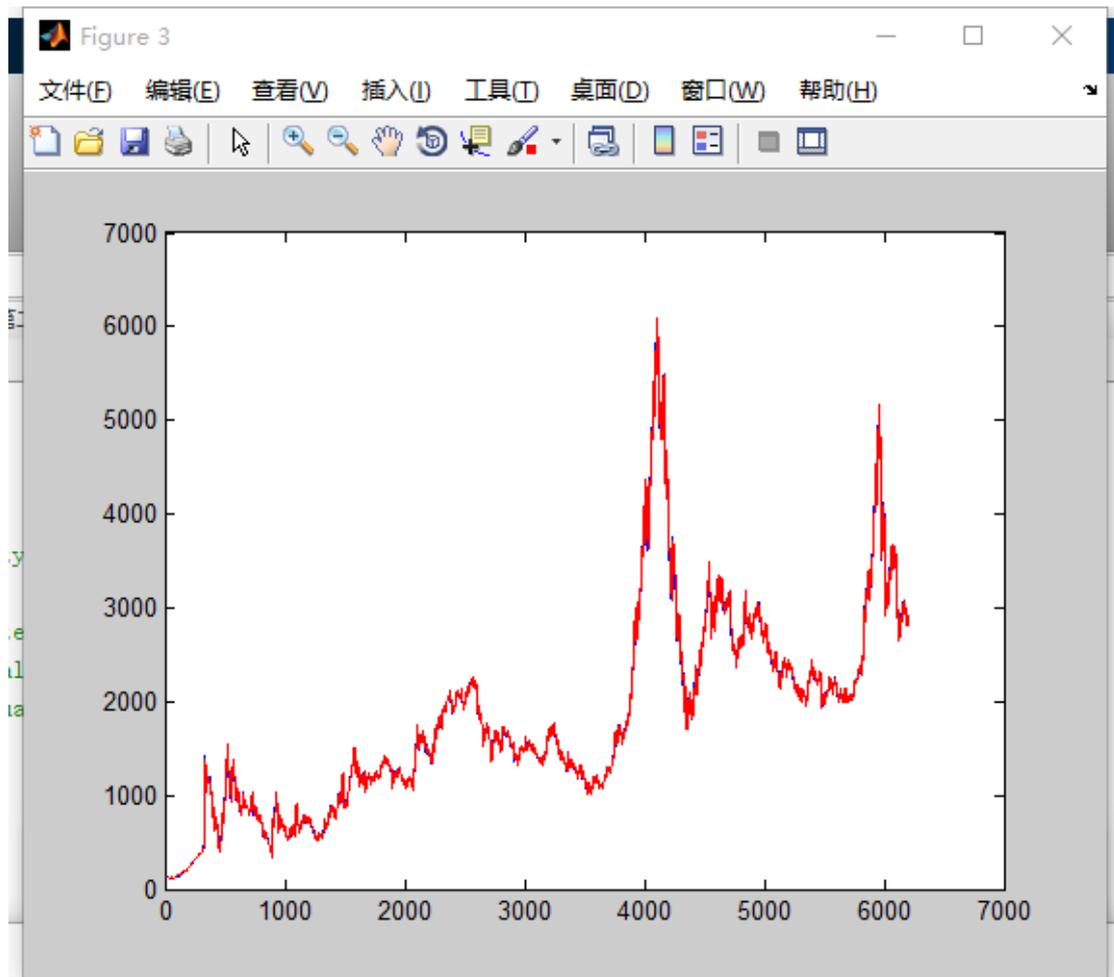
2、输出：

第二天收盘价

3、隐含层个数：

5

4、拟合结果：



0.01 以内的误差准确率: 0.517419

0.02 以内的误差准确率: 0.768710

0.05 以内的误差准确率: 0.955968

涨跌准确率: 0.891975

F.

1、输入指标:

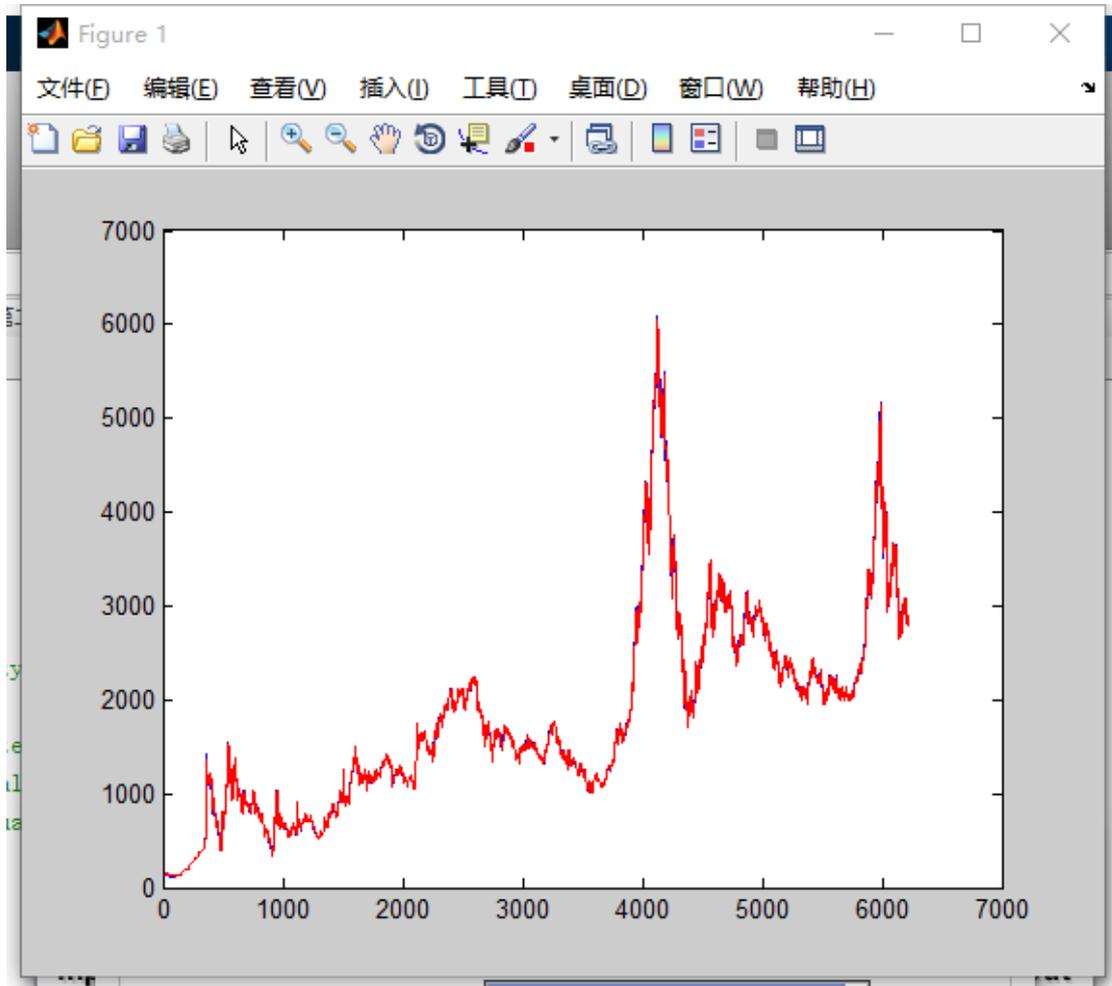
MACD、WMS、RSI、OBV、收盘价

2、输出:

第二天收盘价

3、隐含层个数:

4、拟合结果：



0.01 以内的误差准确率： 0.512289

0.02 以内的误差准确率： 0.764980

0.05 以内的误差准确率： 0.945703

涨跌准确率： 0.885815

4.3 神经网络时间序列模型 Nonlinear Autoregressive (NAR)

4.3.1 神经网络配置

由于采用时间序列模型，故输出层节点个数为 1，而选择滞后值 $d=5$ ，即输入层节点个数为 5，隐含层节点个数为 10；

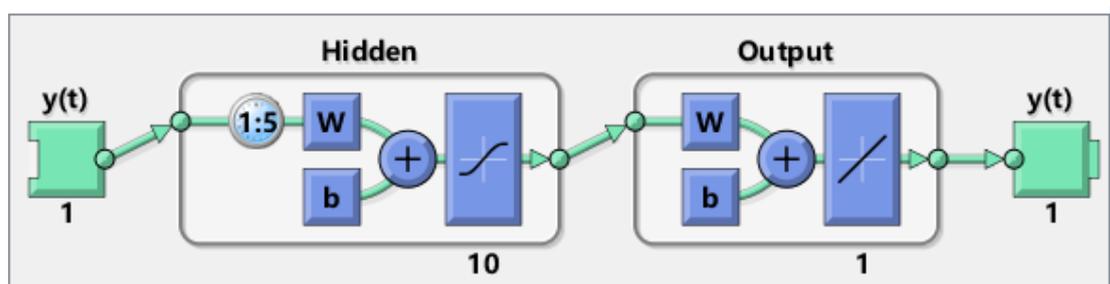
激活函数隐含层选用逻辑回归，输出层选择线性回归；

性能评价选用 MSE；

训练方法选用 Levenberg-marquardt，一种非线性最小二乘法；

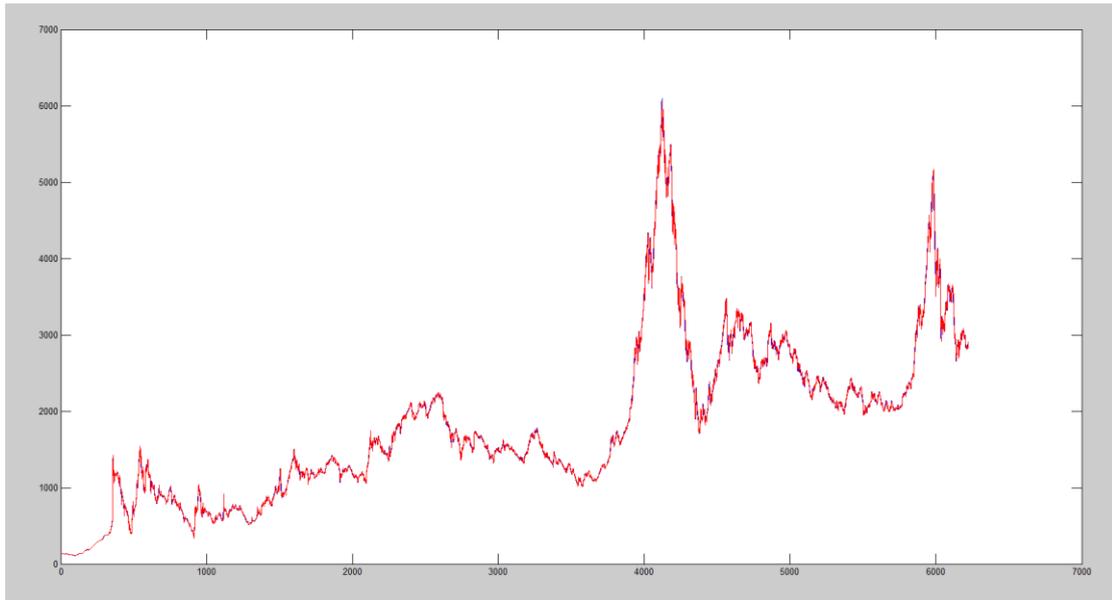
指定训练集为 80%，验证集和测试集均为 10%

由以上选择配置，得到神经网络如下：



4.3.2 数据拟合结果

(蓝色表示实际值，红色表示预测值)



0.01 以内的误差准确率: 0.543000

0.02 以内的误差准确率: 0.782189

0.05 以内的误差准确率: 0.962064

涨跌准确率: 0.521222

4.4 ARIMA 模型

4.4.1 模型选择

使用 R 语言的 `forecast` 包中的 `auto.arima` 选择 ARIMA 模型的参数。

```
> arimaModel<-auto.arima(shangzheng,trace=T)
```

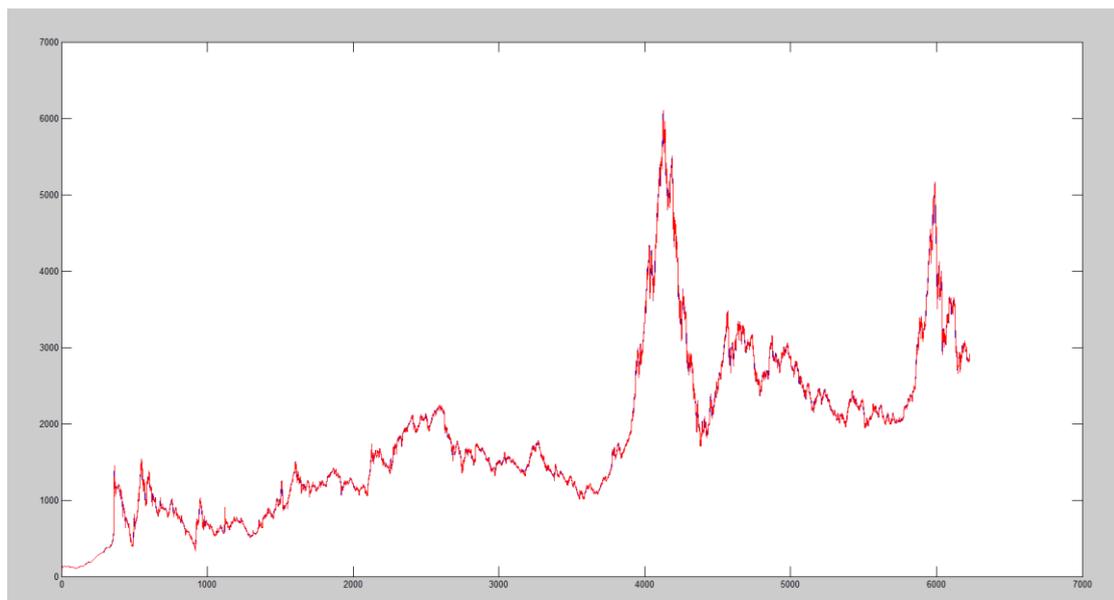
```
ARIMA(2,1,2) with drift : 64250.9  
ARIMA(0,1,0) with drift : 64294.07  
ARIMA(1,1,0) with drift : 64288.01  
ARIMA(0,1,1) with drift : 64286.22  
ARIMA(0,1,0) : 64292.78  
ARIMA(1,1,2) with drift : 64280.18  
ARIMA(3,1,2) with drift : 64229.74  
ARIMA(3,1,1) with drift : 64257.12  
ARIMA(3,1,3) with drift : 64225.52  
ARIMA(4,1,4) with drift : 64227.2  
ARIMA(3,1,3) : 64224.29  
ARIMA(2,1,3) : 64244.87  
ARIMA(4,1,3) : 64227.79  
ARIMA(3,1,2) : 64228.4  
ARIMA(3,1,4) : 64226.81  
ARIMA(2,1,2) : 64249.63  
ARIMA(4,1,4) : 64225.7
```

```
Best model: ARIMA(3,1,3)
```

最终选定模型是 ARIMA (3,1,3) 即自回归 3 阶，滑动平均 3 阶，差分一次。

4.4.2 数据拟合结果

(蓝色表示实际值，红色表示预测值)



0.01 以内的误差准确率: 0.564568

0.02 以内的误差准确率：0.793607

0.05 以内的误差准确率：0.961934

涨跌准确率：0.534137

5. 总结和展望

整个实验拟合结果比较满意，从图中可以看出预测值和真实值相差不大

不同模型结果不同，时间序列预测模型结果更准，但涨跌率并不高，而神经网络拟合模型涨跌率较高，猜测原因 时间序列模型直接利用结果值来预测结果值更直接，同时神经网络拟合模型利用更多的指标，内容更丰富，趋势更准确。

同时可以预测更符合实际需求的参数，比如 7 日平均等来提高准确率。