568Homework3

bf289

February 21, 2019

There are 5 stock data of different companies used to test my stock prediction program: "AAPL", "AMD", "AMZN", "FB", "GOOG", and they are all downloaded from https://www.nasdaq.com

The data format is csv file as following:

	date	close	volume	open	high	low
2	2/19/19	1627.58	3671821	1601	1634	1600.56
3	2/15/19	1607.95	4333397	1627.86	1628.9099	1604.5
4	2/14/19	1622.65	4106555	1624.5	1637.9	1606.061
5	2/13/19	1640	3545096	1647	1656.38	1637.1063
6	2/12/19	1638.01	4830487	1604	1639.395	1598.88
7	2/11/19	1591	3301240	1600.98	1609.29	1586
8	2/8/19	1588.22	5649888	1586	1588.59	1566.756
9	2/7/19	1614.37	4605293	1625	1625.5389	1592.91
10	2/6/19	1640.26	3930870	1670.75	1672.26	1633.34
11	2/5/19	1658.81	4443632	1643.34	1665.26	1642.5
12	2/4/19	1633.31	4911319	1623	1649.63	1613.5
13	2/1/19	1626.23	11484130	1638.88	1673.06	1622.01
14	1/31/19	1718.73	10304330	1692.85	1736.411	1679.0821
15	1/30/19	1670.43	5717042	1623	1676.95	1619.68
16	1/29/19	1593.88	4591586	1631.27	1632.38	1590.72
17	1/28/19	1637.89	4807564	1643.59	1645	1614.09
18	1/25/19	1670.57	4933434	1670.5	1683.48	1661.6122
19	1/24/19	1654.93	4077737	1641.07	1657.2572	1631.7847
20	1/23/19	1640.02	5192420	1656	1657.43	1612
21	1/22/19	1632.17	6402655	1681	1681.8668	1610.2

Figure 1: data

And each file contains the latest one-year historical data for the certain company.

From Feb 19, 2018 to Feb 19, 2019

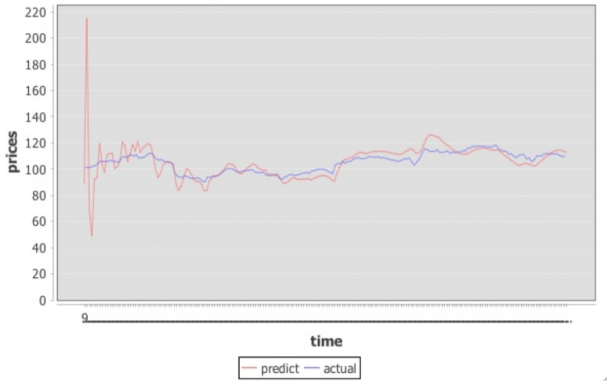

Figure 2: prediction

The Figure 2 shows that when I choose AMZN's historical data to test, and we set the number of input to 200, this means we will predict of 201th day. And the Figure 2 directly shows the differences between actual prices and predict prices from 9 to 200. Because if the number of input is too small, the predict price is absolutely not accurate. And after testing the data, we find that when the number of input is at least 9, the prediciton is good enough to check.

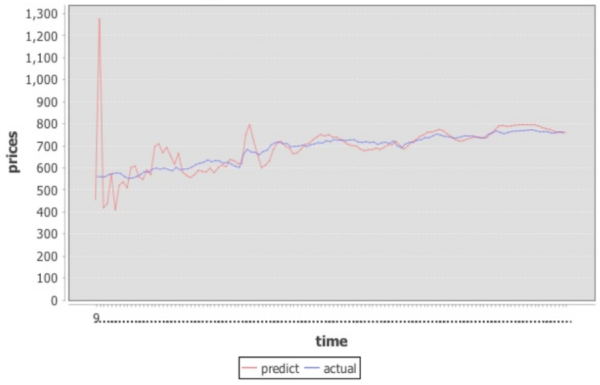

Figure 3: prediction

And Figure 3 shows another predication for FB.

And now, let's talk about the performance evaluation.

And here we will use two different the number of input, 100 and 200, we could clearly see the prediction errors in the following.

MeanErr: absolute mean error

RelErr: average relative error

	N	MeanErr	RelErr	prediction	actual	variation
AAPL	100	6.3982	0.0641	93.43	99.83	0.2383
AMD	100	9.7294	0.0985	89.08	98.81	0.2383
AMZN	100	7.6379	0.0103	728.43	736.07	0.2383
FB	100	1.0135	0.0086	118.36	119.37	0.2383
GOOG	100	29.2722	0.03989	704.51	733.78	0.2383

	N	MeanErr	RelErr	prediction	actual	variation
AAPL	200	0.7980	0.0072	111.83	111.02	0.1374
AMD	200	2.3460	0.0187	123.04	125.39	0.1374
AMZN	200	23.9599	0.0311	794.38	704.19	0.1374
FB	200	2.4944	0.0211	120.44	117.95	0.1374
GOOG	200	2.8119	0.0036	768.38	771.19	0.1374

And we could find that with the increment of the number of input, the variation is decreasing, which means the prediction is more accurate.

And now, we focused on one company, and we utilize different number of input, and check the performance evaluation.

AMZN

N	MeanErr	RelErr
10	380.8983	0.5622
40	72.6193	0.1262
70	44.1934	0.0767
100	32.3187	0.0439
130	26.6389	0.0380
160	22.3931	0.0293
190	14.0385	0.0188
220	8.3391	0.0101

And from the table, we could find, with the increment of N, the errors is going to be smaller, and when N is very small, the prediction is absolutely inaccuracy.