VE 445 Course Project

Project Introduction

This project is meant to practice your ability of handling actual business problems about financial transaction. There is only one task for you to predict the rise and fall of the price, and you should apply as many as models and algorithms as you can including supervised learning and unsupervised learning.

This is a group project, preferably 3 people per group. At the end of the report (Appendix section), you should describe the contributions of each member in the group. After you decide your group members and leader, please add a group in "People" section of Canvas, invite your teammates.

The due date is May 2nd 11:59 PM. You should submit a report for the whole project and your executable source code (you should explain the configuration of your environment and key dependence libraries in report).

Project Report

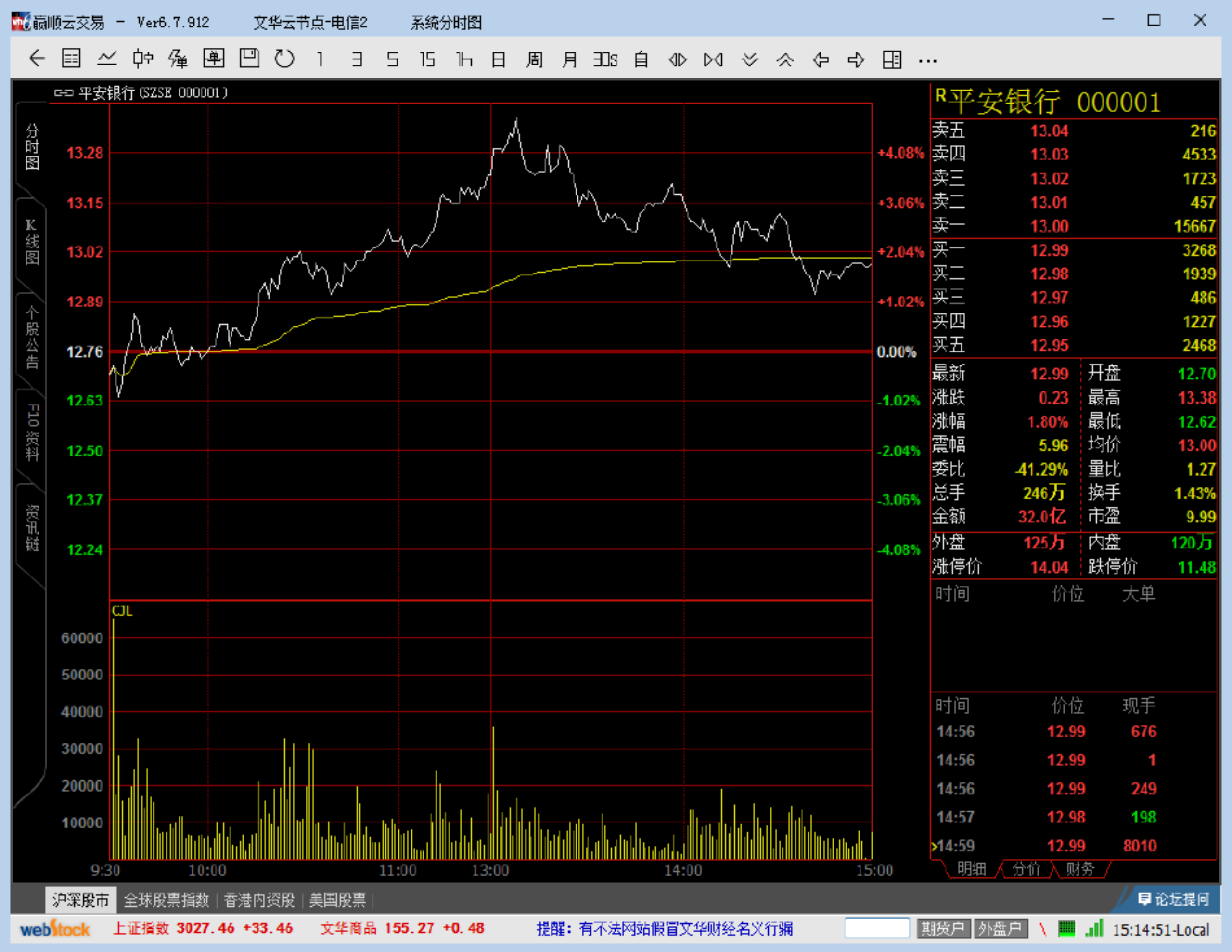
In this lab, the developing environment is not limited to python and tensorflow, you can also use other tools including MATLAB, Mathematica, R language, etc. In the final project report, you should describe your work clearly, including description of the environment configuration, solution, and test results must be included. In environment configuration part, you should describe the software language, key libraries, and other things necessary to run the program. In solution part, you should explain what kind of problems are encountered and how to solve them. In test results part, you should display the actual results of the program and analyze them. Additionally, the contribution of the team members should be described in the appendix section of the report.

Project Background

The business problems of this assignment are derived from financial transactions. Financial transactions are the process of converting asset classes in the market. The purpose of financial transactions is to preserve and increase the value of assets, but improper transactions will suffer loss. There is a class of assets that naturally hold income, such as deposits and bonds. This type of assets is not involved in this project. What is considered in this project is those assets which has no income during the holding period, the profits and losses are only caused by price changes, and trading is the only transaction method. To conclude, the transaction in this project is an attempt to "buy low and sell high".

"Buying low and selling high" is challenging. Both low and high are relative to the future. In essence, this problem needs to grasp the regularity of the market and predict the future. However, the question whether there is regularity in market is in a long-term debate among market researchers. Among them, the famous claims include technical analysis theory, value investment theory, random walk theory, modern asset portfolio theory, effective market hypothesis and behavioral finance. Some claims are completely affirmative, such as technical analysis. Some are completely negative, such as random walks. Even with empirical evidence, there are still many different opinions. Most active fund income cannot exceed passive fund income, but there also exist active fund maintaining excess returns for a long time. This project has a relatively positive view of this issue. All three tasks are trying to find a statistically significant method, making unbiased judgments, and buying and selling.

In order to facilitate the understanding of data and tasks for students who have not touched the transaction, we will briefly introduce the market structure and trading operations in the following part. Students who are familiar with transaction can skip this part.



R平安银行 000001

卖五	13.04	216
卖四	13.03	4533
卖三	13.02	1723
卖二	13.01	457
卖一	13.00	15667
买一	12.99	3268
买二	12.98	1939
买三	12.97	486
买四	12.96	1227
买五	12.95	2468

最新	12.99	开盘	12.70
涨跌	0.23	最高	13.38
涨幅	1.80%	最低	12.62
震幅	5.96	均价	13.00
委比	-41.29%	量比	1.27
总手	246万	换手	1.43%
金额	32.0亿	市盈	9.99
外盘	125万	内盘	120万
涨停价	14.04	跌停价	11.48

时间	价位	大单

时间	价位	现手
14:56	12.99	676
14:56	12.99	1
14:56	12.99	249
14:57	12.98	198
>14:59	12.99	8010

明细 分价 财务

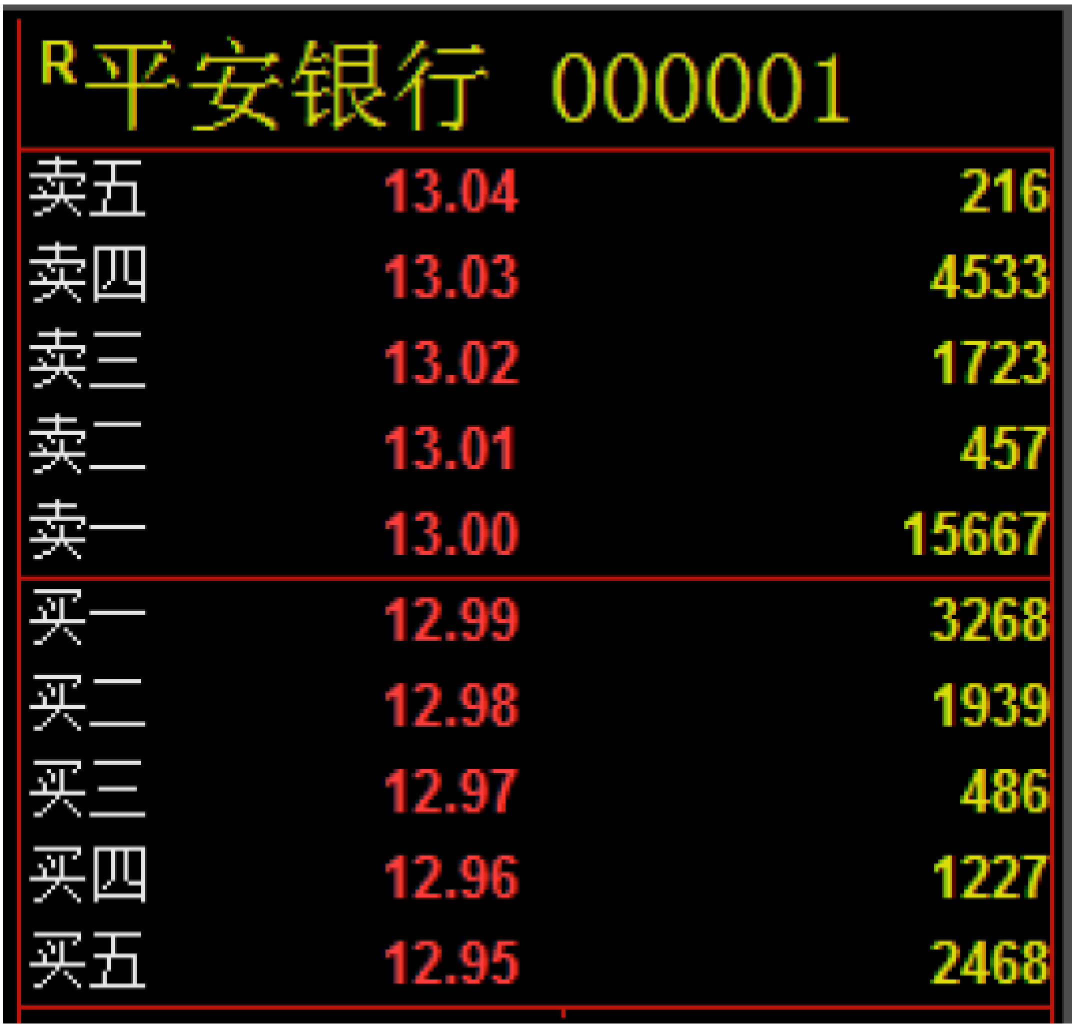
上证指数 3027.46 +33.46　文华商品 155.27 +0.48

提醒：有不法网站假冒文华财经名义行骗

Fig.1 Typical Software Trading Interface

Fig.1 is a typical software trading interface showing market trading information. The white line in the middle of the picture is the historical information of the price per minute, and the yellow bar below is the trading volume per minute. The upper right part is the unfilled declaration order at the current moment. The middle right part is the statistical information of the market. The lower right part is the information of the latest deal order.

During the market opening period, continuous bidding is adopted, and the transaction rules are generally “price priority, time priority”. The specific measures are as follows: for each sale and purchase order declared, if the purchase price is greater than or equal to the sale price of the unfilled declaration, or the sale price is less than or equal to the purchase price of the unfilled declaration, then the sale price or purchase price is already reported. Otherwise, it can become an unfilled declaration order waiting for a transaction, also known as a pending order. Pending orders can be maintained throughout the trading day.



R平安银行 000001

卖五	13.04	216
卖四	13.03	4533
卖三	13.02	1723
卖二	13.01	457
卖一	13.00	15667
买一	12.99	3268
买二	12.98	1939
买三	12.97	486
买四	12.96	1227
买五	12.95	2468

Fig.2 Unfilled Pending Order

We can take Fig.2 as an example to specify the order commission. In the figure, "sell one (卖一)" to "sell five (卖五)" refers to the five cheapest price of sold pending orders, and each price is one stall. The first column of data refers to the price of the sold pending order of each stall, the unit is the yuan, the second column of data is the number of sold pending orders of each stall, the unit is the hand. "buy one (买一)" to "buy five (买五)" refers to the five most expensive purchased pending orders, others are similar. If a new purchase is made at this time, the price is 12.99 and the quantity is 1000. Since "sell one" is 13.00, no transaction for the purchase, but the number of "buy one" will rise to 4268. If a new purchase is made at this time, the price is 13.00 and the quantity is 1000. Then the buy-in will be sold, the trading volume is 1000 (different exchanges may be different, some exchanges calculate bilaterally, that is, the volume is 2000), and the number of "sell one" will drop to 14667. If a new purchase is made at this time, the price is 13.01 and the quantity is 1000. Then the buy-in will be sold, but the transaction price is 13.00, the number of "sell one" will drop to 14667, but the "sell two" will not change. If a new purchase is made at this time, the price is 13.00 and the quantity is 20000. Then the purchase will be partially sold, the volume will be 15667, and the stall will change. The original "sell two" becomes the current "sell one", the price and quantity remain unchanged. The original "sell three" become the current "sell two", the price and quantity remain unchanged, others are similar. The current "buy one" will have price 13.00 and volume 4333. The original "buy one" becomes the current "buy two", the price and quantity remain unchanged. The original "sell three" become the current "sell two", the price and quantity remain unchanged, others are similar.

When buying in, the direction of selling is called the counterpart, and vice versa. After the purchase is made, the position is increased, and after the sale is made, the position is reduced. When the position number is positive, it holds long positions, and when the position number is negative, it holds short positions.

When the market transaction is closed, the handling fee is removed. If the average price of purchased orders is less than the sold orders, there exists profits, and this is the ultimate goal of the following tasks of this project. But the following datasets and tasks will package and simplify the specific financial data and business, making it a pure mathematical problem. When solving problems, you may not have extensive experience in financial business, but follow the task guidelines. But if you want to achieve better results, you need to understand the data and logic behind it.

Dataset

The dataset includes the raw data and the provided features. The raw data is the public data pushed by the market, and each tick pushes a group. The following data is provided:

1. InstrumentID (contract name)
2. TradingDay (transaction date)
3. UpdateTime (update time-stamp)
4. UpdateMillisec (updated millisecond time-stamp)
5. LastPrice (the price of the last transaction)
6. Volume (accumulated volume during the trading day)
7. LastVolume (volume after the last push of data)
8. Turnover (accumulated transaction amount during the trading day)
9. LastTurnover (the amount of the transaction after the last push of the data)
10. AskPrice5 (price for selling stall 5)
11. AskPrice4 (price for selling stall 4)
12. AskPrice3 (price for selling stall 3)
13. AskPrice2 (price for selling stall 2)
14. AskPrice1 (price for selling stall 1)
15. BidPrice1 (price for buying stall 1)
16. BidPrice2 (price for buying stall 2)
17. BidPrice3 (price for buying stall 3)
18. BidPrice4 (price for buying stall 4)
19. BidPrice5 (price for buying stall 5)
20. AskVolume5 (volume of 5 selling stall 5)
21. AskVolume4 (volume of 5 selling stall 4)
22. AskVolume3 (volume of 5 selling stall 3)
23. AskVolume2 (volume of 5 selling stall 2)
24. AskVolume1 (volume of 5 selling stall 1)
25. BidVolume1 (volume of 5 buying stall 1)
26. BidVolume2 (volume of 5 buying stall 2)
27. BidVolume3 (volume of 5 buying stall 3)
28. BidVolume4 (volume of 5 buying stall 4)
29. BidVolume5 (volume of 5 buying stall 5)
30. OpenInterest (Contract opening quantity)
31. UpperLimitPrice (limit-up price)
32. LowerLimitPrice (limit-down price)

Environment Configuration

It is recommended to perform tasks in Python environment. You should install necessary libraries and process the datasets. Try to use parallel computing, whether you use CPU or GPU.

All tasks do not limit the models and methods to be used. For beginners, you can start with the installation of anaconda and build a basic Python environment. At the same time, scikit-learn is installed by default in anaconda, which has basic models and algorithms. Numpy and Pandas are mainstream tools for algorithm implementation and data processing. Numpy supports accelerated vector operations, which is more efficient than using a 'for' loop in Python. If you want to try a neural network, you can configure tensorflow, which has a CPU version and a GPU version. There are also many environment management tools, 'conda' on anaconda, git on github and pip.

It is recommended that you do not implement the algorithm yourself if you have a ready-made algorithm implementation and library. Implementing algorithms is not the main task of this project.

Task 1. Prediction on Price Rise and Fall

This task involves supervise learning. The basic requirement is to learn a certain model and do classification about the price rise and fall according to the features and labels in the dataset.

The price rise and fall after n time stamps at time t is defined as:

$$L = \mathrm{sgn}(\mathrm{AskPrice1}_{t+n} + \mathrm{BidPrice1}_{t+n} - \mathrm{AskPrice1}_t - \mathrm{BidPrice1}_t)$$

$$\mathrm{sgn}(x) = \begin{cases} 1 & x < 0 \\ -1 & x \geq 0 \end{cases}$$

where the value of n can be customized. In the end, you should report your methods' final results, with the prediction accuracy on the testing set.

This project will be evaluated according to the difference between your predicted results and true values. Tail value will be assigned with higher weight (you should make accurate prediction in the long run). Complex models not necessarily guarantee better results, because financial transaction problems depend on the problem itself and the size of the dataset.

Here are a few typical models and algorithms for reference:

1. SVM
2. BP networks
3. CNN
4. RNN
5. K-means
6. Logistic Regression.
7. Naive Bayes Method.

Here are a few typical models and algorithms for reference:

1. Deep Direct Reinforcement Learning for Financial Signal Representation and Trading.
2. Agent Inspired Trading Using Recurrent Reinforcement Learning and LSTM Neural Networks.

Submission

1. This is a group work and you should team up in 3-person group voluntarily. After you have teamed up, you should sign in the group of VE445/People/VE445 Project. Only one submission for each group.
2. The due is May 2nd 11:59 PM, no due extension.