Columbia University
MATH GR5260 Spring 2022
Programming for Quant and Computational Finance
Ka Yi Ng

Homework Assignment 4
Release date: April 15, 2023 (Sat)
Due date: April 30, 2023 (Sun) 11:00pm

HOMEWORK GUIDELINE

Submit your solution file(s) onto CourseWorks by the specified due date and time.

A. Theory parts

    Your solution to theory questions can be (i) hand-written and scanned as a pdf file or (ii) prepared using Word and be converted into a pdf file or (iii) be included in the solution for practice parts.

B. Practice parts

    Your Python source code and outputs shall be prepared and submitted in a format (eg. pdf, html, .ipynb, .py, etc.) that the graders can understand. Before submitting your solution, make sure that you have run your program successfully to generate all required outputs. Points may be deducted if some outputs are missing.

    You are free to use standard python packages distributed by Anaconda.

    Do not submit additional packages that can be downloaded from the web. Just provide instructions in the solution file.

C. Other files

    Include files that may be requested by the homework assignment. Make proper references in the solution file.

Note: If you suspect there are typos in this homework, or some questions are wrong, please first discuss with your TAs.

QUESTION 1

Many researchers and practitioners have looked into using various machine learning models and techniques in stock price prediction based on different sets of data (eg. technical indicators, micro and macro-economic indicators, etc.).

In this question, we will use the time series of the daily log price returns of the following stocks as input features to train an SVM model to predict the direction of the next day's return of DJI index.

- AAPL, HD, JNJ, JPM, MSFT, UNH, V, XOM

a) Use pandas_datareader to download from Yahoo Finance the adjusted closing prices of the tickers required for the period from Jan 2nd 2019 to Dec 31st 2020 (both inclusive).

b) (Data preparation) Prepare training dataset and test dataset using the following guidelines.

- Identify any missing values in the downloaded data. Fill in the missing values using the previous trading day's value. For example, if a ticker value on 2019-07-03 is missing, fill it with the its value on the previous trading day, 2019-07-02.

- Use the first 80% (starting from Jan 4th 2019) as the training dataset and reserve the remaining rows to form the test dataset. The labels are defined as the sign of the <u>next day's</u> return of DJI index. That is, if the daily log price return on a trading day is positive, the sign of the daily return is 1; otherwise, it is -1.

- Rescale the feature values using standardization (ie. StandardScaler).

c) Use the prepared training dataset in part (b) to fit an SVM model. Use RBF kernel with $\gamma = 1, C = 1$.

d) Use 5-split time series cross validation (ie. TimeSeriesSplit) to compute the mean of the cross-validated scores for the trained model.

e) Use GridSearchCV to search for the best hyperparameters C, γ over $C = 0.1, 1.0, 10, 100$ and $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5, 1, 5, 10$.

We would simulate a simple trading strategy with the trained SVM model and see how much profit or loss would be attained over the period from Oct 1st 2020 to Dec 31st 2020 (both inclusive).

The simple trading strategy is to use the predicted signals to buy all-in or sell all-out. At most one trade would be executed at the start of each trading day based on the predicted trend for the day's return.

Assume that there is an index tracking fund that perfectly mirrors the movements of DJI index. Any fractional amount of the fund can be traded without any transaction costs. Also assume that at the start of each trading day, shares can be purchased or sold at the closing price on the previous day.

Starting Oct 1st 2020, observe the predicted signal for the next day's return. If the predicted signal for next day's return is a buy, then invest the full capital amount in the index fund at the market start of the next day; otherwise, do nothing and wait till the first day that a predicted signal is a buy, invest the full capital amount in the index fund. After the investment, hold the shares until the first day that the predicted signal is a sell, then sell all the shares. Repeat this trading strategy till the end of the period. For example, if the predicted signals for day 1 to day 6 are: -1, 1, 1, -1, -1, 1, then the trade actions are: day 1: do nothing, day 2: buy all-in, day 3: hold, day 4: sell all-out, day5: no nothing, day 6: buy all-in.

f) Suppose the initial capital amount is $10000. Compute the final amount at the market close of Dec 31st 2020 after applying the above trading strategy over the specified period.

g) Consider the simplest buy-and-hold strategy: at the market start of Oct 1st 2020, invest the full initial capital amount in the fund and hold the shares till the end of Dec 31st 2020. Compute the final amount at the market close of Dec 31st 2020. Comment and compare results from these two strategies.

--end of Assignment—