Indian Institute of Technology Bombay; August 24, 2023
EE782 Advanced Topics in Machine Learning

Assignment 1: LSTM-based Stock Trading System

Instructions:

- Assignment has to be done individually, in Google CoLab using python and pytorch
- Write copious comments using text cells and in the code cells to explain your intent and demonstrate understanding
- End with a cell that lists sources of starter code (e.g., github repo links, or chatGPT) for all modules
- Submit the (1) ipynb file, (2) its py version, (3) link to a 10 minute video explaining the code, your intent, and your observations and decisions on Moodle
- CODE WITHOUT COMMENTS FOR EACH LINE OF CODE WILL NOT BE GRADED
- VIDEO WITHOUT VIEW PERMISSIONS WILL NOT BE GRADED. Test your video by sending link to a friend
- Deadline is Sept 10, 2023, 11:59pm

1. Familiarize yourself with the input data sp500_tickers_A-D_1min_1pppix.zip: [1]
    a) Plot the minute-by-minute closing price series of few stocks
    b) Plot the day-by-day closing price series of a few stocks
    c) Plot a complete candlestick chart with volume on secondary y-axis for a few stocks with a time period of your choice
    d) Note down your observations, e.g. are there any data issues, unexpected jumps, unexpected missing data etc.
2. Try at least two ways to normalize the data as stock prices and volumes change over time and are different across companies, and pick one by noting your justification. [1]
3. Make some scenario decisions: [1]
    a) high-frequency trading or intra-day swing, or inter-day trade, or long-term (multi-day or multi-week or multi-month).
    b) Assume a buy-ask spread (inversely related to volume and directly related to price) and trade commission based on a quick market research. Your trade will lose the buy-ask spread and commissions every time you trade.
    c) Decide if you will trade only one stock, or have a model to trade a basket from a particular industry, or any stock.
4. Write a pytorch module for defining an LSTM model. Keep it flexible so that the input dimension, number of units, number of layers can easily be changed. [1]
5. Write a flexible dataloader for training the LSTM model, especially if you are high frequency data. The inputs should be open, close, high, low, volume of one or more stocks (e.g. other stocks that can help predict the chosen stock price). [2]
6. Train or pre-train the model by trying to predict the future price (or change in price, if normalized) and keep the future horizon flexible (e.g., easy to change between one minute or 10 minutes into the future). You can try to predict the opening or closing time. Leave the last two years out for testing. [2]
7. Set up a trading module that will make some hard-coded logical decisions to buy, hold, sell (in any order, because in shorting you can sell first and buy later). [2]
8. Test the trading system on the latest years on which the model was not trained. [3]
    a) Does the price prediction error increase as you go further from the last time on which it was trained?
    b) Can you profitably trade with the bid-ask spread and commissions taken into account?
    c) How does your profitability compare to a simple buy-and-hold strategy over long term (e.g. one or two years)?
9. Advanced: [2 bonus]
    a) Can you now modify the model to use multiple stock prices as inputs to predict a single stock (your choice)? Does it improve predictions?
    b) Can you add day of the week, day in year, and time as inputs? Does it improve results?