

Mini-Project #9

Due by 11am on Wednesday, June 7th

Instructions

- You can work in groups of up to four students. If you work in a group, please submit one assignment via Gradescope (with all group members' names).
- Detailed submission instructions can be found on the course website (<https://web.stanford.edu/class/cs168>) under “Coursework - Assignments” section.
- Use 12pt or higher font for your writeup.
- Make sure the plots you submit are easy to read at a normal zoom level.
- If you’ve written code to solve a certain part of a problem, or if the part explicitly asks you to implement an algorithm, you must also include the code in your pdf submission.
- Code marked as Deliverable should be pasted into the relevant section. Keep variable names consistent with those used in the problem statement, and with general conventions. No need to include import statements and other scaffolding, if it is clear from context. Use the `verbatim` or `minted` environments to paste code in L^AT_EX.

```
def example():  
    print "Your code should be formatted like this."
```

- **Reminder:** No late assignments will be accepted, but we will drop your lowest assignment grade.

Part 1: Gaming the Stock Market with Multiplicative Weights?

Goal: The multiplicative weights algorithm is often described as a “game” of choosing stocks to invest in each day, with a theoretical guarantee comparing the performance of your investments to that of the single best stock, in hindsight. So...to what extent does multiplicative weights actually work as an investment strategy? [Disclaimer: this assignment is *not* investment advice!!!]

Description: In this problem, you will work with daily stock data from 489 stocks included in the S&P 500 index fund. Suppose you are given a budget of \$1 to invest in the stock market each day. We’ll explore how the multiplicative weights strategy performs on stock data, where we view the i th stock as the i th “expert” and its loss on day t is

$$-\frac{C_{t+1}(i) - C_t(i)}{C_t(i)},$$

where $C_t(i)$ denotes the close price of stock i on day t . This loss can be interpreted as the amount of money lost on day t if you invested \$1 in stock i —equivalently, this loss is the negative of the amount you would have earned by investing \$1 in that stock).

The file “close.csv” contains the daily closing price of 489 stocks in the S&P. The file “tickers.csv” contains the ticker symbol of each of the 489 stocks. The i th column of close.csv should be a length 1259 vector representing the close prices for the i th stock in “tickers.csv” for each day the stock market is open from Jan 1 2018 to Jan 1 2023.

- (a) (1 point) First, let's check to see how well the best stock did (since our theoretical guarantees will be in terms of this). Load the data from the file “close.csv”. Calculate the return achieved by the best stock over the 5 years of data, and plot the cumulative return of this stock over the 5 years (so your X-axis should be 5 years). Note that you should be computing the earnings if you invest \$1 each day in that stock, *not* the total earnings from investing \$1 at the beginning of the 5 years.

For comparison, calculate the return achieved by investing in a portfolio that puts an equal weight on each stock (ie investing 1/489 dollars in each of the stocks each day), and plot on the same axes the return of the equal weight portfolio.

- (b) (8 points) Using Theorem 5.1 (and Remark 5.2) from Lecture 17, compute the return *guaranteed* by the multiplicative weights algorithm. You will need to plug in the maximum loss value and the number of days, and then compute the optimal choice of ϵ . What does this mean—how optimistic should we be that multiplicative weights will give a reasonable investment strategy?

- (c) (6 points) One more question before actually running multiplicative weights on the stock data. Suppose the return of each of the 489 stocks, on each of the 1259 days, were drawn independently from a Normal distribution of mean 0, and standard deviation 0.01 (corresponding to the stock having increased or decreased by roughly 1%). Empirically, what is the total return of investing \$1 each day in the single best stock (chosen in hindsight)? Plot your earnings of this investment over the 1259 days. In this setting with independent, zero mean returns, is it possible for *any* investment strategy (that doesn't already know the future) to have positive expected earnings?

- (d) (8 points) Run the multiplicative weights (MW) algorithm to choose an investment portfolio for each day for the following choices of ϵ : [0, 0.01, 0.1, 1, 2, 4, 8]. For each ϵ , plot the cumulative return achieved by MW over the course of the 5 years (have all the plots on the same axes). What investment strategy does $\epsilon = 0$ correspond to? What is the total return for each ϵ , and which choice of ϵ does best?

- (e) (3 points) Investing in a non-diverse portfolio can be risky. First, let's look at how many stocks are well-represented in the MW at each day in the 5 years. One measure of this diversity is the maximum probability that the distribution assigns to any one stock (if the maximum is small—say < 0.1 , then the distribution is diverse...). For each value of ϵ , plot this maximum probability for each of the days (so the x-axis of the plot will go from 0 to 1257). For $\epsilon \geq 2$, what is this investment strategy actually doing during the last 2 or 3 years of the investment period? (Which stocks is it investing in? Does it invest in a number of different stocks or not?)

- (f) (2 points) In general, one often evaluates an investment strategy as some sort of ratio of the return, to the volatility. One simple metric is the ratio of the returns to the standard deviation in the daily returns. For each value of ϵ , what is the ratio of the average daily return to the standard deviation in the daily returns for that ϵ ? Do the larger returns still seem to be better by this risk-adjusted metric?

- (g) (7 points) Suppose we want our investment strategy to have a “recency” bias, in the sense that we would like to discount a stock's performance in the distant past when calculating the weight. (For example, if a stock performed extremely well 4 years ago, but has been flat since then, we would not want to put much weight on it.) Run the a modified version of multiplicative weights, where you weighted the historical returns of each stock so that the returns t days ago are multiplied by a factor of α^t for some $\alpha < 1$. Consider $\alpha \in [0.98, 0.99, 0.995]$ which corresponds to basing the weights on roughly the past 50, 100, and 200 days, respectively.

Repeat the parts (d), (e), and (f) for these values of α . Which value of ϵ lead to the best returns? What choices of ϵ lead to the best ratio between return and standard deviation for each α ? For $\epsilon \geq 2$, how often does the stock with the highest weight change, and how does this compare with your answer in part (e)?

- (h) (8 points) So, how robust is the performance of multiplicative weights? Let's do a single stress-test: zero out the returns of stock ENPH—the single best-in-hindsight stock, and redo parts d and g. Think about what you see and discuss in a few sentences.

- (i) (6 points) Based on these experiments, what do you think of MW as a strategy for investing stocks? Discuss in a few sentences.
- (j) (1 point) Please acknowledge the following disclaimer: this miniproject should not be construed as investment advice.