Homework 8:
Multiplicative weights

The stock market is often described as a "game" of choosing stocks to invest in each day, where your goal is to match the performance of the best performing stock. This is literally a regret minimization problem...right? So...to what extent does multiplicative weights actually work as an investment strategy?

Disclaimer: this assignment is absolutely NOT actual investment advice! We are not responsible if you lose all of your money!!!

In this problem, you will work with daily stock data from 419 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 ith stock as the ith "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 fractional 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 these 419 stocks. The file `tickers.csv` contains the ticker symbol of each of the 419 stocks. The ith column of `close.csv` should be a length 1259 vector representing the close prices for the ith stock in `tickers.csv` for each day the stock market is open from May 28th, 2019 to May 24th, 2024.

Problem 1: Setup and Theoretical Analysis

(a) [1 point] Load the data from `close.csv`. Compute and plot the cumulative return of the best-performing stock over the five years. Compare it with an equal-weight portfolio investing $\frac{1}{419}$ in each stock daily.

(b) [4 points] Discuss, from a theoretical perspective, why it makes sense to take the normalized difference in cost as the loss function, as opposed to the unnormalized loss $-(C_{t+1}(i) - C_t(i))$.

(c) [8 points] Using what we covered in lecture, compute the return guarantee for the multiplicative weights algorithm. Determine the theoretically optimal choice of ε and discuss expectations.

(d) [6 points] Simulate the scenario where all stock returns are randomly and independently either 1 or 0 every day, for all companies. Compute and plot the cumulative return of the best stock in hindsight. Discuss whether any strategy can achieve positive expected earnings in this setting.

(e) [4 points] What does $\varepsilon = 0$ represent? On the other hand, what is the limiting behavior of the algorithm as we take $\varepsilon \to \infty$?

Problem 2: Multiplicative weights for stock data

(a) [6 points] Implement the multiplicative weights algorithm with $\varepsilon \in \{0, 0.01, 0.1, 1, 2, 4, 8\}$. Plot cumulative returns for each on the same axes. Identify the best-performing ε. Does this correspond with the theoretically optimal ε that you determined in Problem 1? Can you give a plausible reason why it does or does not correspond?

(b) [3 points] A basic principle of investing is that it's risky to not diversify. One basic measure of diversity is the maximum probability that the algorithm assigns to any stock. For each value of ε, plot this maximum probability that multiplicative weights assigns for each of the days. For $\varepsilon \geq 2$, what is our investment strategy actually doing during most of the investment period? (Which stocks is it investing in? Does it invest in a number of different stocks or not?)

(c) [8 points] Now, let's do a bit of a stress test. Remove the best performing stock (hint: this should be a certain GPU company...) by zeroing out its corresponding column. Now redo parts (a) and (b). Discuss the results you see in a couple of sentences. Do you see any qualitative differences in behavior?

(d) [6 points] Based on your experiments, assess the viability of multiplicative weights for investing. Discuss in a few sentences the results you found.

(e) [4 points] One somewhat unappealing aspect (amongst many others...) of muliplicative weights for this setting is that in a certain sense, it treats all losses equally over time. Explain what this means more concretely, and propose a method that puts more weight on recent data. You don't have to run experiments with this method.

(f) [1 point] Acknowledge that this homework is **not** investment advice. Please don't recreate this homework with real money.

(EC) [0-4 points] That being said...can you come up with any alterations or modifications to multiplicative weights that perform better on both the vanilla dataset as well as the dataset with the best peforming stock removed? Explain what you did, and report the total reward your method achieves on both datasets.