Reinforcement Learning and Dynamic Optimization

Assignment 2 - Day Trading in the Stock Market

Manoudaki Chressida 2019030201

May 15, 2024

1 Introduction

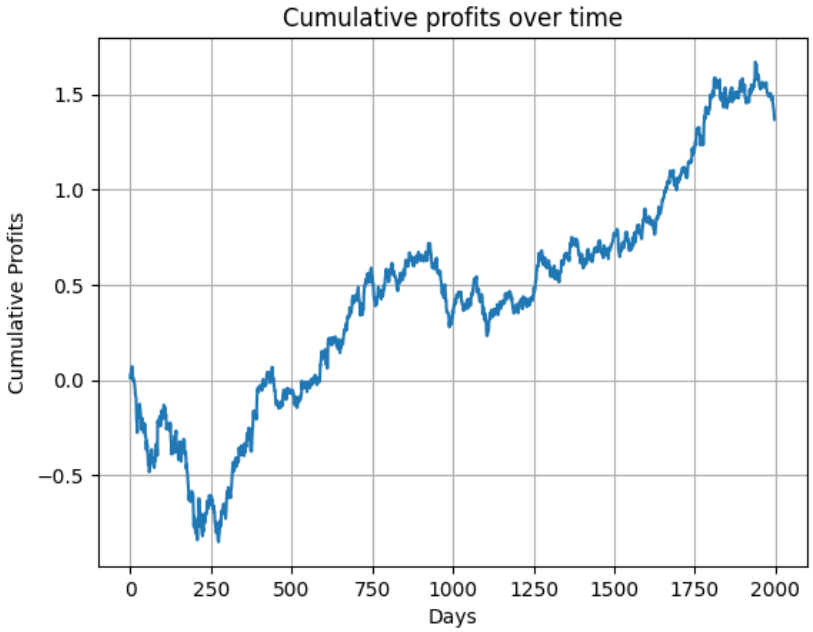
In this assignment, the adversarial framework of bandit algorithms was used to optimize stock investments in a simplified day-to-day trading scenario. The goal was to maximize profits by applying bandit algorithms, including experts, adversarial bandit algorithms, and Online Convex Optimization (OCO). The setup simulates a simplified trading environment where 1 euro is invested each day in one of K stocks, aiming to optimize cumulative profit over 2000 days.

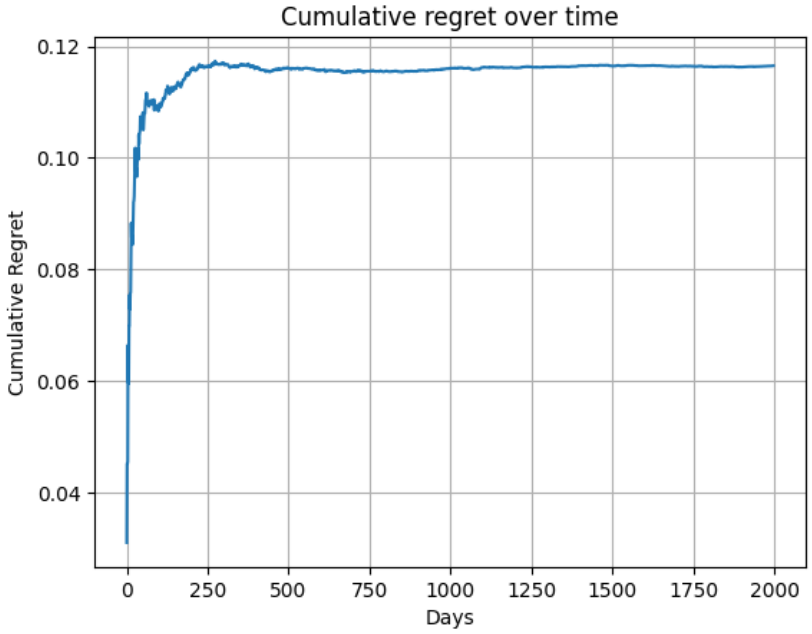
2 Tasks

Three main tasks were designed to test different aspects of bandit algorithms in stock trading.

2.1 Task 1: Experts Setup

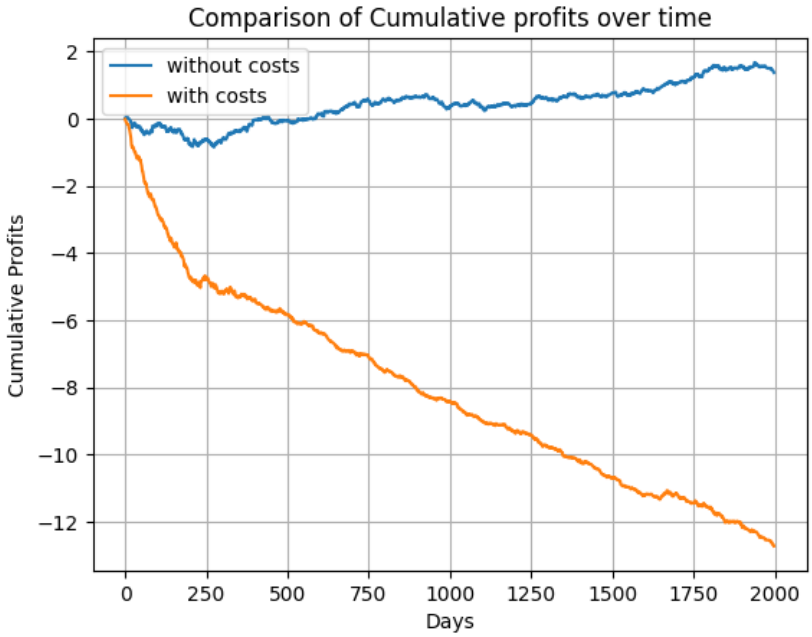
The Multiplicative Weights algorithm was implemented with full feedback, where the price change percentages for all K stocks are learned at the end of each day. Below are the plots illustrating the cumulative profit and cumulative regret of the algorithm from the first day to the last.

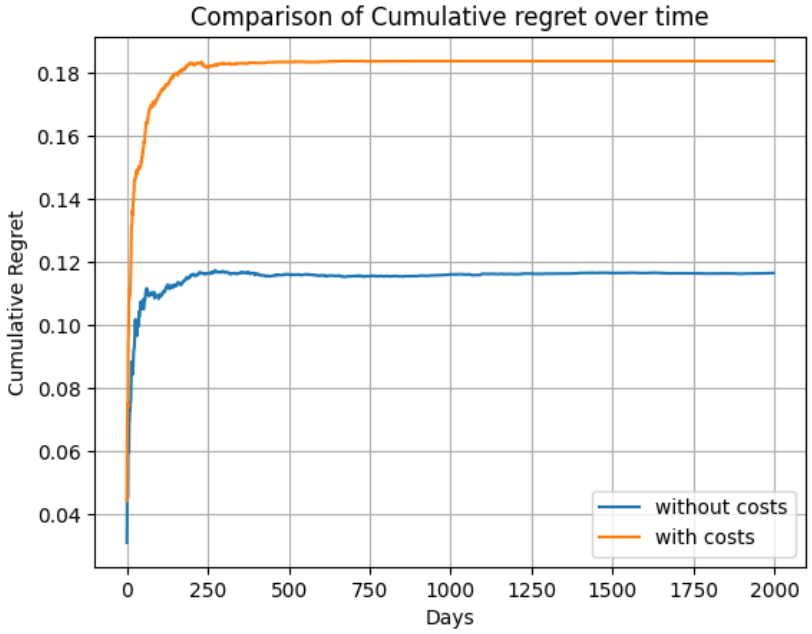




2.2 Task 2: Experts with Transaction Fees

The algorithm was modified to deduct transaction costs from daily gains. The weights were updated to reflect these costs, and the cumulative regret and profit plots showed the impact of transaction fees on overall performance. Below are the plots illustrating the cumulative profit and cumulative regret of the algorithms from Task 1 and Task 2 from the first day to the last.

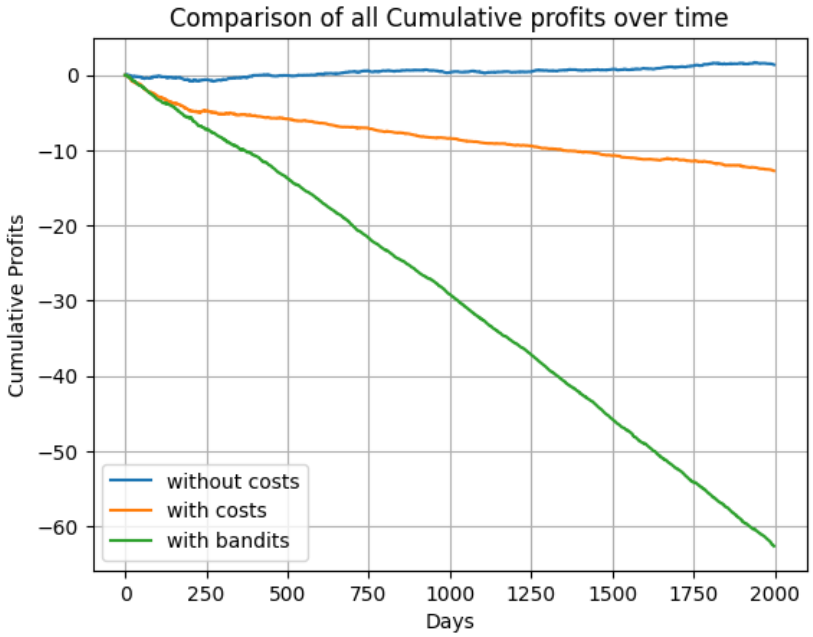


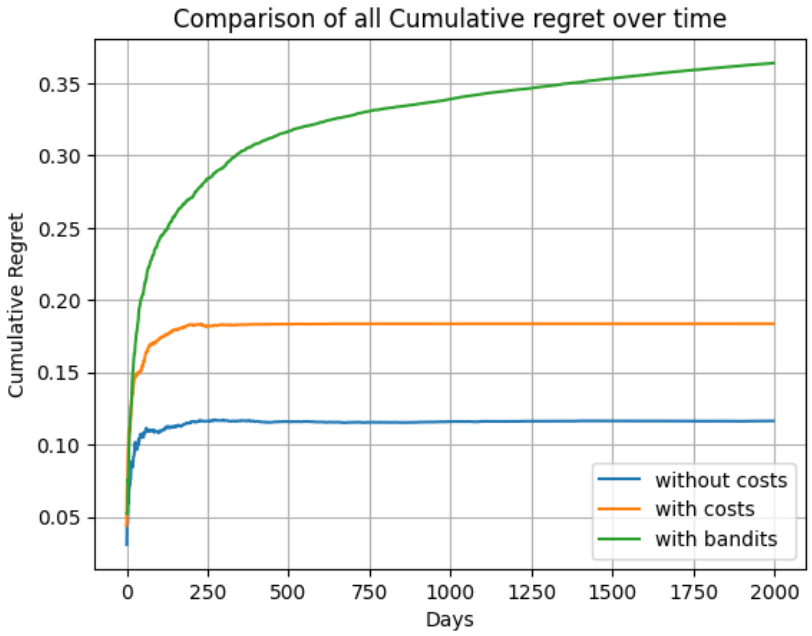


Including Transaction fees in Task 2 resulted in a noticeable reduction in cumulative profits, showing the impact of such costs on trading strategies. These fees, add extra costs to each trade, making it harder to see how well each stock is really performing. This extra cost usually increases the regret compared to Task 1 because it makes the trading environment more complex and noisy. In other words, the added transaction fees make it more challenging to pick the best stocks, leading to higher regret.

2.3 Task 3: Bandits with Transaction Fees

The bandit setup required modifications to handle limited feedback. An exploration-exploitation trade-off was used, and results were compared with the experts setup to evaluate performance under restricted information conditions. Below are the plots illustrating the cumulative profit and cumulative regret of the algorithms from Task 1, Task 2 and Task 3 from the first day to the last.





In the bandit setting, you only receive feedback on the chosen stock's performance each day. This partial feedback makes it harder to learn the optimal strategy because the algorithm must balance exploration (trying out different stocks) and exploitation (investing in stocks that seem promising based on limited feedback). The inclusion of transaction costs further complicates the learning process. Thus, higher regret compared to the full feedback scenario of the other tasks, is observed. This limited information makes it much harder to make informed investment decisions, leading to less effective trading and lower overall profits. Additionally, the inclusion of transaction fees in Task 3 further reduces the cumulative profits, as these costs must be subtracted from any gains made.