Yijie Bai

DSCC435: Optimization for Machine Learning

November 7, 2023

Course Project Proposal

Convex Optimization Techniques for NIFTY-50 Portfolio Optimization

Machine learning has revolutionized the way we approach problem-solving across various domains, particularly in the financial sector. At the core of its transformative power lies the ability to optimize predictive models, enabling the extraction of meaningful patterns from complex datasets. Convex optimization [1], a subset of optimization techniques, has proven particularly effective in refining these models. It focuses on the challenge of minimizing convex functions, which, due to their unique properties, guarantee global solutions to optimization problems, making them highly suitable for applications in machine learning.

In financial markets, where the stakes are high and the data is intricate, the application of machine learning through Portfolio Optimization has become a critical endeavor. It represents a confluence of data science and financial expertise, aimed at constructing investment portfolios with optimal risk-reward profiles. By leveraging convex optimization algorithms, financial analysts can devise strategies that maximize return while controlling for market volatility. This synergy of finance and machine learning not only enhances the decision-making process but also paves the way for more resilient investment frameworks.

The NIFTY-50 Stocks Dataset presents a fertile ground for applying these sophisticated optimization techniques. Comprising the stock prices of the leading 50 companies in the National Stock Exchange of India, it embodies the intricate dynamics of a burgeoning financial market. The dataset's diversity in company size, sector, and volatility offers a comprehensive canvas to test the efficacy of various convex optimization methods taught in contemporary machine learning courses.

This project proposal aims to undertake a thorough exploration of convex optimization algorithms, as delineated in the academic curriculum, and their application to the NIFTY-50 dataset. By doing so, it seeks to bridge theoretical knowledge with practical financial analysis, providing a robust validation of machine learning methods in the context of real-world financial data. Through this endeavor, the project will not only contribute to the academic discourse but also provide actionable insights for the finance industry.

Current Outline:

Problem Statement:

Portfolio optimization is a critical task that seeks to maximize return for a given level of risk, or alternatively, to minimize risk for a given level of expected return. Despite the prevalence of traditional methods like Markowitz's Modern Portfolio Theory, the dynamic and often non-linear nature of financial markets necessitates more advanced approaches. Machine learning, with its ability to parse complex datasets and uncover non-intuitive patterns, can offer significant improvements. However, choosing the right optimization algorithm is paramount to success. This project will investigate the performance of various machine learning optimization algorithms on the NIFTY-50 Stocks Dataset to determine the most effective methods for portfolio optimization.

Data Collection:

Our primary dataset will be the NIFTY-50 Stocks Dataset from Kaggle [2]. This dataset includes historical stock prices, volumes, and other relevant financial indicators necessary for building robust portfolio optimization models. We may enrich this dataset with necessary additional data points in the future such as macroeconomic indicators, market sentiment analysis from news articles and financial reports, aiming to create a comprehensive dataset for our optimization task.

Methodologies and algorithms:

- Data Collection: Utilizing the NIFTY-50 Stocks Dataset available on Kaggle, which includes price data of the top 50 stocks in the Indian market.
- Data Cleaning: Processing this dataset to handle missing values, outliers, and any corrupt records to ensure quality and reliability.
- Exploratory Data Analysis (EDA): Analyzing the dataset to identify trends, patterns, and correlations that could influence portfolio optimization strategies. Feature Importance Analysis: By employing various analytical techniques, we'll identify the most influential factors shaping public opinion.
- Machine Learning Optimazation Algorithms: This project will implement a series of machine learning optimization algorithms, such as Gradient Methods, Subgradient Methods, Proximal Gradient Method, Accelerated Gradient Methods, and Smoothing Techniques. Each algorithm will be tailored to the specific characteristics of portfolio optimization problems. We will explore both traditional convex optimization algorithms and more recent advancements like Accelerated Gradient Methods and Proximal Gradient Methods, which are particularly effective for high-dimensional and sparse optimization problems.
- Other possible algorithms: we may use other algorithms such as Markowitz Model [3].

Performance Evaluation:

- Return on Investment (ROI): Measuring the profitability of the optimized portfolios.
- Volatility: Evaluating the risk associated with the optimized portfolios through their volatility.
- Sharpe Ratio: Calculating the risk-adjusted return to determine how well the return of an asset compensates the investor for the risk taken.
- Computational Efficiency: Considering the time complexity and resource usage of each algorithm, which is crucial for processing large datasets and for real-time trading scenarios.
- Robustness: Testing the optimized portfolios against market downturns and unexpected market conditions to ensure they are not overly fitted to the historical data.

Reference:

[1]. Stanford EE364a: Convex Optimization (http://web.stanford.edu/class/ee364a/lectures.html)

[2]. NIFTY-50 Stocks Dataset(https://www.kaggle.com/datasets/iamsouravbanerjee/nifty50-stocks-dataset)

[3]. Portfolio Optimization using Markowitz Model(https://www.coursera.org/projects/portfolio-optimization-markowitz-model)