Stock Portfolio Optimization Using Genetic Algorithm

Kemal Bayik

Abstract

In this report, stock portfolio optimization was performed using Genetic Algorithm (GA). 15 stocks were selected from the S&P 500 and a portfolio containing all or some of these stocks was created by using the prices of these stocks between May 5, 2012 and December 31, 2022. The purpose of the GA is to optimize the weight of these stocks in the portfolio by maximizing the sharpe ratio. The portfolio was tested with stock prices between January 1, 2023 and January 1, 2024 and its percentage return was calculated. The results obtained from this experiment provided a return approximately 2.55 times higher than the S&P 500 index, and it can be said that this was a successful result. In addition, optimized portfolios were created with different crossover rate, mutation rate and population size values and the effect of these parameters on the convergence time was observed. According to the result, changing the crossover rate and population size parameters has almost no effect, while reducing the mutation rate increases the convergence time to the optimal solution.

1. Introduction

Stocks are one of the important investment instruments today. Many people make some of their investments in stocks and aim to get a return above inflation. Investors generally create a portfolio of many stocks instead of buying a single stock and aim to minimize risk in this way. The goal of this project is to create a portfolio consisting of some or all of the 15 stocks in the S&P500 using Genetic Algorithm (GA) [1]. GA tries to maximize the sharpe ratio [2] by optimizing the weights of these stocks. Using the price data[9] of the stocks between May 20, 2012 and December 31, 2022, the GA will optimize the stock weights in the portfolio and calculate the percentage return of the created portfolio from January 1, 2023 to January 1, 2024. The performance of the GA will be evaluated by comparing it with the return of the S&P500 index between January 1, 2023 and January 1, 2024.

GAs were first introduced by Holland, 1976[1] and have been applied to many optimization problems in the past and successful results have been obtained. Portfolio optimization discussed in this report is one of these problems. In Sefiane, Benbouziane, 2012[3], a portfolio consisting of 5 assets was optimized using GA. At the same time, single point, two point and arithmetic crossover operators are compared. Xiaolou Yang, 2006[4], introduced a GA that significantly increases the accuracy of asset return estimation and expected return estimation by using both past information and future uncertainty.

Stock portfolio management is an example of adaptive systems. The portfolio must be able to adapt to changing market conditions, risk tolerance and investor preferences. So, adaptive systems in the context of portfolio optimization can be defined as follows. Adaptive systems, in the context of portfolio optimization, are systems that can evolve according to market conditions, risk tolerance and investor preferences, and try to maximize the rate of return and minimize the rate of risk.[5][6]

2. Methods

In this section of the report, the GA, problem formulation, experimental setup and used libraries are explained in detail.

2.1. Genetic Algorithm

In the GA used in the project, a single objective (sharpe ratio) was tried to be optimized. In this section, the GA is explained in more detail under 4 subsections. The pseudocode of the algorithm is shown in Algorithm 1.

Algorithm 1 Genetic Algorithm

```
Initialize parameters
Generate population with random individuals
for each generation do
    Calculate fitness for each individual in the population
    Add the two best individuals to the new population
    Select parents from the population with tournament selection
    Generate new population through arithmetic crossover and mutation
    Assign new population to the population variable
end for
return best individual from the final population
```

2.1.1. Selection

Tournament selection was used as the selection method in the algorithm. This method organizes a tournament among 5 randomly selected individuals from the population and selects the individual with the highest fitness value from this group. The selected individual is used as the parent for crossover and mutation operations. In addition, elitism was used in the algorithm and the two best individuals were added to the new population.

2.1.2. Mutation

The purpose of mutation is to prevent premature convergence and maintain diversity in the population. Mutation was implemented in the algorithm as follows. If a randomly selected number between 0 and 1 is less than the initially determined mutation rate value, a random number from the Gaussian distribution is added to the weight of a random stock in the portfolio. The genes of the individual are then normalized to ensure that the sum of the portfolio weights is 1, that is, to ensure that the portfolio weights form a valid distribution.

2.1.3. Crossover

The crossover function simulates the biological crossover. This process is too important for introducing new solutions into the population. Arithmetic crossover is used in the GA. The reason for this is that, in a previous study on portfolio optimization with GAs,[3] it was seen that the arithmetic crossover operator gave better results than other crossover operators compared. In the GA, crossover works as follows. First of all, a number between 0 and 1 is randomly selected, and if this number is less than the crossover rate value determined at the beginning, the crossover process is performed. In the crossover process, an alpha(α) value is chosen randomly. This value determines how the parents' genes will be combined in new individuals. The formula for calculating the genes of the first child is shown in Formula (1), and the formula for calculating the genes of the second child is shown in Formula (2). Afterwards, the normalization process is applied here as well.

$$\text{child}_1[i] = \alpha \cdot \text{parent}_1[i] + (1-\alpha) \cdot \text{parent}_2[i] \tag{1}$$

$$\text{child}_2[i] = \alpha \cdot \text{parent}_2[i] + (1-\alpha) \cdot \text{parent}_1[i] \tag{2}$$

2.1.4. Fitness

The fitness function is used to measure how "fit" an individual is. Sharpe ratio was used as a fitness function in the algorithm. Sharpe ratio shows how much risk is taken against the potential return of the portfolio and is calculated with the formula shown in Formula (3). R_p in the formula represents the expected return of the portfolio. The expected return of the portfolio is calculated with the historical average method and its formula is shown in Formula (4). The Historical Average Method estimates the expected return by averaging the past returns of the portfolio. In Formula (4), N represents the number of years and $R_{p,t}$ represents the portfolio return at time t. R_f in Formula (3) represents the risk-free interest rate. In the algorithm, 1 year treasury yield, that is, the yield of 1 year maturity Treasury bills issued by the United States Government, was used as the risk-free interest rate. The price data between 20/05/2012 and 30/12/2022 was averaged and used as a constant in the formula. The reason why 1 year treasury yield is used as the risk-free interest rate is that treasury bills are considered risk-free investment options because they are backed by the payment guarantee of the US government. σ_p in Formula (3) represents the standard deviation of the return of the portfolio, that is, the volatility of the portfolio. The formula for the volatility of the portfolio is shown in Formula (5). In this formula, w represents the vector containing the portfolio weights., Σ represents the covariance matrix of returns, and 252 represents the business days in a year.

$$S = \frac{R_p - R_f}{\sigma_p} \tag{3}$$

$$R_p = \frac{1}{N}\sum_{t=1}^{N} R_{p,t} \tag{4}$$

$$\sigma_p = \sqrt{w \cdot ((\Sigma \times 252) \cdot w)} \tag{5}$$

2.2. Problem Formulation

Decision Variables:

- Let w_i represent the weight of asset i in the portfolio, for $i = 1, 2, \ldots, n$, where n is the number of assets in the portfolio.

Parameters:

- R_p: The expected return of the portfolio.
- σ_p: The volatility of the portfolio.
- R_f: The risk-free rate of return.

Objective:

- The objective is to maximize the Sharpe Ratio of the portfolio, defined as:

$$S = \frac{R_p - R_f}{\sigma_p} \tag{6}$$

Constraints:

1. The sum of the weights must equal 1, ensuring full allocation of the portfolio:

$$\sum_{i=1}^{n} w_i = 1 \tag{7}$$

2. Weights must be non-negative:

$$w_i \geq 0 \quad \forall i \tag{8}$$

2.3. Experimental Setup

In this section, the experimental setup to measure the success of the GA is introduced.

The coding of the experiment was done with Python. Compiled in Visual Studio Code and run on Macbook Pro with M1 chip, 32GB RAM configuration.

The algorithm was run for 100 iterations with these parameters. The results obtained and shown are the average of the results obtained in these 100 iterations. At the same time, the algorithm was run with different crossover rate, mutation rate and population size values and the effects of these parameters were tried to be measured. In order to measure its effect accurately, when a parameter was changed, the other parameters were kept constant at their initial values.

For example;

- Crossover rate = 0.7, Mutation rate = 0.7, Pop. size = 100
- Crossover rate = 0.3, Mutation rate = 0.7, Pop. size = 100

The initial values of the crossover rate, mutation rate and population size variables are 0.7, 0.7, and 100.

Selected stocks:

- AAPL
- MSFT
- GOOG
- AMZN
- NVDA
- META
- LLY
- AVGO
- TSLA
- JPM
- UNH
- V
- XOM
- MA
- JNJ

Parameters:

- Population size : 100 & 300
- Number of generations : 200
- Number of tests : 100
- Tournament size : 5
- Mutation rate : 0.7 & 0.3
- Crossover rate : 0.7 & 0.3
- Elitism size : 2
- Risk free rate : % 0.9524

2.4. Used Libraries

In this section, the Python libraries used in the project are introduced.

- numpy (version 1.26.3)
- pandas (version 2.2.1)
- yfinance (version 0.2.37)
- matplotlib (version 3.8.2)
- random

3. Results and Analyses

In this section, the results obtained from the experiment are introduced. The resulting optimized portfolio was obtained by taking the average of the stock weight values of 100 optimized portfolios obtained in 100 iterations. The data presented in all graphs or tables except Figure 2 are obtained from the average values of 100 optimized portfolios. The weights of the stocks in the optimized portfolio are shown in Figure 1 and Table 1, and the returns of 100 optimized portfolios on the test data are shown in Figure 2.

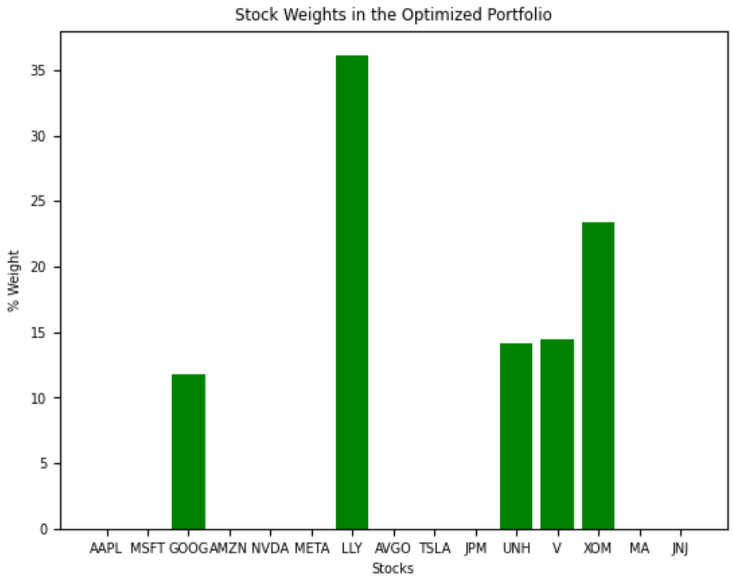

Figure 1: Stock Weights in the Optimized Portfolio

Stock Weights in the Optimized Portfolio	
Stock	% Weight
AAPL	0
MSFT	0
GOOG	11.82
AMZN	0
NVDA	0
META	0
LLY	36.15
AVGO	0
TSLA	0
JPM	0
UNH	14.13
V	14.49
XOM	23.41
MA	0
JNJ	0

Table 1: Stock Weights in the Optimized Portfolio

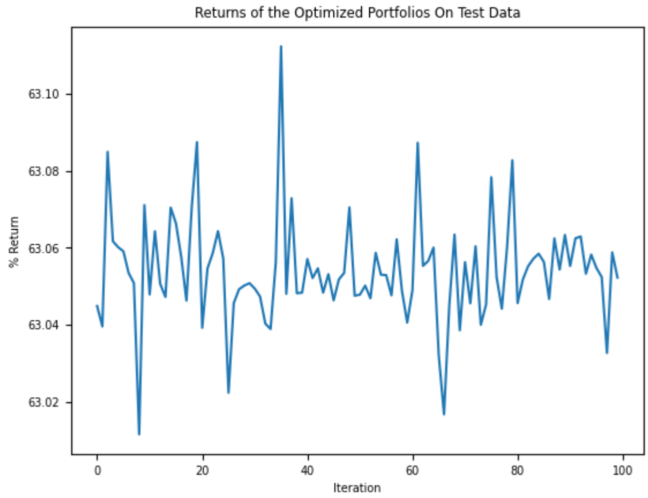

Figure 2: Returns of the 100 Optimized Portfolios on Test Data

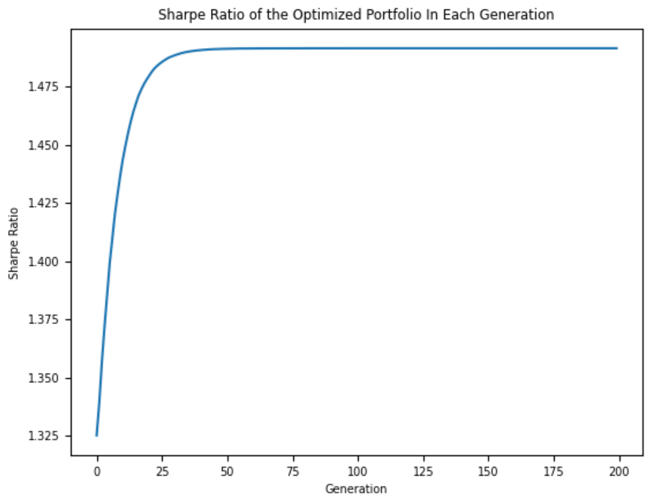

Figure 3: Sharpe Ratio of the Optimized Portfolio In Each Generation

As seen in the Figure 3, the sharpe ratios of the optimized portfolio increase in each generation and creating an upward trend. From this we can conclude that GA can adapt and finds better solutions over generations. As the generation progresses, we see that the rate of increase of the sharp ratio slows down, the graph becomes more horizontal and the GA converges to the optimal solution. However, we can say that GA reaches the optimal solution approximately in the 45th - 50th generations and does not show any improvement in later generations. From here it can be concluded that if the generation number of the GA is reduced from 200 to 50 - 60 generations, the GA will reach the optimal solution and time and working power can be saved. Finally, the sharpe ratio of the optimized portfolio is approximately 1.5 in the last generation. In the financial sector, a sharpe ratio between 0 and 1 is interpreted as bad, between 1 and 1.99 as good, between 2 and 2.99 as very good, and above 3 as excellent [10][11][12]. Then, it can be said that the optimized portfolio gives good result.

As shown in Figure 4 and Table 2, the optimized portfolio provided a 63.05% return between 01/01/2023 and 01/01/2024, while the S&P500 index provided a 24.73% return. The portfolio's increase of 63.05 percent in a time period when the index increased by 24.73 percent can easily be considered a successful result.

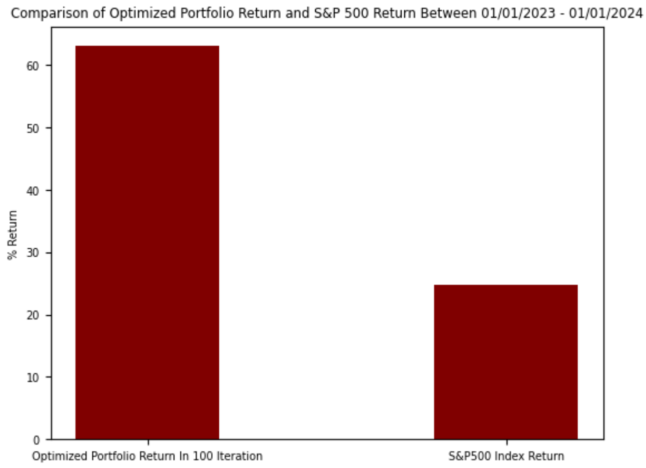

Figure 4: Comparison of the Optimized Portfolio Return and the S&P 500 Return Between 01/01/2023 and 01/01/2024

Comparison of the Optimized Portfolio Return and the S&P 500 Return Between 01/01/2023 and 01/01/2024	
Optimized Portfolio Return	S&P 500 Return
% 63.05	% 24.73

Table 2: Comparison of the Optimized Portfolio Return and the S&P 500 Return Between 01/01/2023 and 01/01/2024

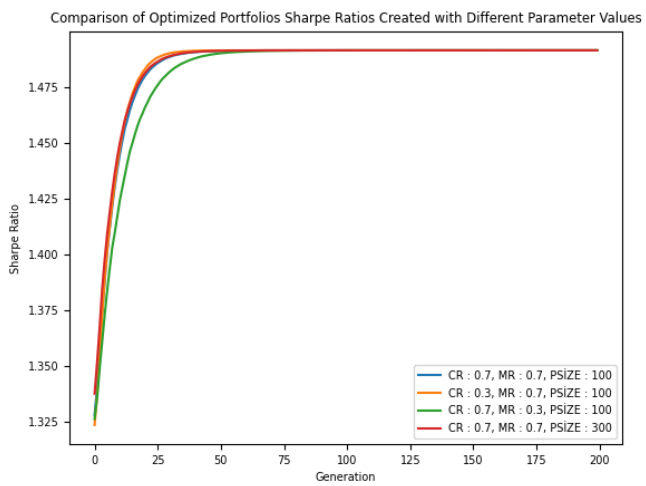

Figure 5: Comparison of optimized portfolios created with different parameter values

Looking at Figure 5, it can be seen that the sharpe ratio values obtained with different parameter values are almost not affected when the crossover rate and population size are changed, but when the mutation rate is reduced, the convergence time increases. However, looking at the final result, all algorithms converged to the same result in approximately the 70th generation.

4. Discussion

According to the definition of adaptability I made in the introduction, the adaptive system I created has some strengths and weaknesses. First of all, I mentioned in the definition that the system should adapt to market conditions. The adaptive system I created using GA can adapt to three market conditions: market prices, market volatility and risk free rate. However, the market conditions mentioned in the definition include economic indicators, interest rates, inflation rates, political and economic news and many other external factors that can affect market prices. Apart from this, the system can adapt according to the market price, volatility and risk free rate and gives good results, as seen in Figure 3 and Figure 4. The other two factors I mentioned in the definition, risk tolerance and investor preferences, do not affect the adaptability of the system.

The contribution of the experiment to adaptive systems is that it shows that GAs can give good results in complex financial problems such as portfolio optimization. The contribution of the experiment to the field of finance is that it shows the importance of adaptability in portfolio management. It is also important to show that GAs work well and are usable in this field.

The system introduced in this report can also be used in risk management and algorithmic trading areas in finance. In addition, it can be used in sectors and projects such as optimization of energy storage systems in the energy sector, optimization of treatment protocols in the healthcare sector and route optimization in logistics.

The system is open to improvement in many aspects. One of these is the inclusion of a few different factors in the market conditions in the definition I made in the introduction section. For example, factors such as interest rate, inflation rate, general market trend, and economic growth data can be included in the system to make the optimized portfolio more reliable against dynamic market conditions. In addition, the investor's preferences and risk tolerance can be included in the system and an optimized portfolio can be created specifically for the investor. Finally, the GA used in this project tries to optimize a single objective, but in addition, other metrics such as sortino ratio and treynor ratio can be added to the problem to create a more comprehensive and reliable portfolio. Algorithms such as NSGA-II[7] and SPEA2[8] can be used for this.

References

[1] J H Holland.Adaptation in Natural Artificial Systems [M]. MIT Press,1976.
[2] Sharpe, W. F. (1966). "Mutual Fund Performance". Journal of Business. 39 (S1): 119– 138.
[3] Sefiane, Slimane and Benbouziane, Mohamed (2012): Portfolio Selection Using Genetic Algorithm. Published in: Journal of Applied Finance & Banking , Vol. 2, No. 4 (2012): pp. 143-154.
[4] Yang, Xiaolou. (2006). Improving Portfolio Efficiency: A Genetic Algorithm Approach. Computational Economics. 28. 1-14. 10.1007/s10614-006-9021-y.
[5] Hommes, Cars. (2001). Financial Markets as Nonlinear Adaptive Evolutionary Systems. Quantitative Finance. 1. 10.1080/713665542.
[6] Hommes, C. H. (2002). Modeling the stylized facts in finance through simple nonlinear adaptive systems. Proceedings of the National Academy of Sciences, 99(suppl_3), 7221-7228.
[7] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." IEEE transactions on evolutionary computation 6.2 (2002): 182-197.
[8] Zitzler, Eckart et al. "SPEA2: Improving the strength pareto evolutionary algorithm." (2001).
[9] https://pypi.org/project/yfinance/
[10] https://www.forbes.com/advisor/investing/sharpe-ratio/
[11] https://www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp#:~:text=Usually%2C%20any%20Sharpe%20ratio%20greater,1.0%20is%20considered%20sub%2Doptimal.
[12] https://www.cmcmarkets.com/en-gb/trading-guides/the-sharpe-ratio-definition