DEEC
DEPARTAMENTO DE ENGENHARIA
ELETROTÉCNICA E DE COMPUTADORES
TÉCNICO LISBOA

Applied Computational Intelligence 2023/2024

Project 2 – EAs for Single and Multi-Objective Optimization

(Week 7)

This project aims at applying Evolutionary Computation methods to solve a **Computational Finance Challenge**. In section 1, the problem to be addressed is briefly introduced. In section 2, the fundamental concepts from Computational Finance required to solve the problem are presented. In section 3, the problem to be solved is detailed described. In section 4, the details on the work submission (code and report) are given. Finally, in section 5, the evaluation process is detailed.

1. Problem Description

The use of Computational Intelligence in the area of Computational Finance is widespread nowadays.
The term stock market refers to several exchanges in which shares of publicly held companies are bought and sold. There are over 60 stock exchanges in the world. Of these, there are 16 exchanges with a market capitalization of $1 trillion or more, and they account for 87% of global market capitalization. By country, the largest stock markets as of January 2022 are in the United States of America (about 59.9%), followed by Japan (about 6.2%) and United Kingdom (about 3.9%).
The stock data analysis and algorithmic trading strategies is specially challenging due to the volatility and unpredictability of markets, however, there are several examples of **applying computation intelligence** to optimize investment strategies. The problem of selecting the optimal point to open and close a position (Long and Short) in stock market is crucial to maximize returns and minimize risks. The large amounts of historical data, from the stock markets, together with a constant feed of new data makes the task of selecting the optimal point to open and close a position extremely complex.
The problem to be solved in this project is the application of Single and Multi-Objective Evolutionary Computation techniques to maximize return and minimize risk, when investing in the stock market using technical indicators. Naturally, the problem is made simple to allow its implementation within the scope of this course. The details of problem to be solved will be detailed in the following sections.

2. Fundamental Concepts

In this section, fundamental concepts, required to solve the proposed Computational Finance problem, will be presented.

2.1. Stock Market Data

The stock data series can be easily obtained from different data sources, such as, [Yahoo Finance], [Google Finance], etc. In this project, we have selected 10 data series from 10 data SP500 stocks, as described in table 3.1. These stock data series include daily stock data for each of the 10 selected stocks as described in table 3.2. Fig. 2.1 illustrates the evolution of the Apple data series along one year, considering daily data for the stock close value.

Fig. 2.1: Apple data series sample.

2.2. Market Positions

The problem to be addressed aims at identifying the optimal open and close positions in stock market. For that purpose, and for simplicity, we will consider just long and short positions in the stock market as described next.

2.2.1. Long Position Definition and Example

A long—or a long position (see fig. 2.2) —refers to the purchase of an asset with the expectation it will increase in value—a bullish attitude. Going long on a stock or bond is the more conventional investing practice in the capital markets, especially for retail investors. With a long-position investment, the investor purchases an asset and owns it with the expectation that the price is going to rise.

2.2.2. Short Position Definition and Example

In finance, being short (see fig. 2.2) in an asset means investing in such a way that the investor will profit if the value of the asset falls. This is the opposite of a conventional "long" position, where the investor will profit if the value of the asset rises. There are a number of ways of achieving a short position. The most fundamental method is "physical" selling short or short-selling, which involves borrowing assets (often securities such as shares or bonds) and selling them. The investor will later purchase the same number of the same type of securities in order to return them to the lender. If the price has fallen in the meantime, the investor will have made a profit equal to the difference. Conversely, if the price has risen then the investor will bear a loss. The short seller must usually pay a fee to borrow the securities (charged at a particular rate over time, like an interest payment), and reimburse the lender for any cash returns such as dividends that were due during the period of lease.

Fig. 2.2: Examples of trading long and short.

2.3. Return-On-Investment (ROI)

Return on investment (ROI) is a ratio between net income (over a period) and investment (costs resulting from an investment of some resources at a point in time). A high ROI means the investment's gains compare favorably to its cost. As a performance measure, ROI is used to evaluate the efficiency of an investment or to compare the efficiencies of several different investments. In economic terms, it is one way of relating profits to capital invested. Finally, ROI does not consider the holding period, and so it can miss opportunity costs of investing elsewhere.

Fig. 2.3: Examples of trading long and short, considering the AAPL data.

2.3.1. ROI from a Long Position

Consider Fig. 2.3 and the buy (entry) and sell (close) points (green points) for a long position. In this case, the ROI is calculated by the following expression:

$$ROI_{LongPosition} = \frac{Sell_{Value} - Buy_{Value}}{Buy_{Value}} x100\%$$

So, for the illustrated case the ROI would be (175-150)/150*100% = 16,7%
Notice that with long positions the investor can obtain an ROI between 0 and ∞.

2.3.2. ROI from a Short Position

Consider Fig. 2.3 and the sell (entry) and buy (close) points (red points) for a long position. In this case, the ROI is calculated by the following expression:

$$ROI_{ShortPosition} = \frac{Sell_{Value} - Buy_{Value}}{Sell_{Value}} x100\%$$

So, for the illustrated case the ROI would be (150-125)/150*100% = 16,7%
Notice that with short positions the investor can obtain an ROI between –∞ and 100%, that is why short selling is considered a high-risk strategy.

2.3.3. Accumulated ROI

For the problem being addressed, for simplicity, consider the initial capital is split in two, half to be used in long positions and half to be used in short positions. So, the achieved total ROI would be given by:

$$ROI_{Total} = \left(ROI_{Long} + ROI_{Short}\right)/2$$

where the ROI_{long} and the ROI_{short} are obtained, respectively by:

$$ROI_{Long} = \sum_i ROI_{Long\_i} \qquad ROI_{Short} = \sum_i ROI_{Short\_i}$$

2.4. Drawdown as Risk Measure

For the single-objective problem the objective will be maximize the ROI, and for the multi-objective problem an investment risk measure must be considered. For the addressed problem, the risk measure will be the drawdown. A drawdown is a peak-to-trough decline during a specific period for an investment, trading account, or fund. A drawdown measures the historical risk of different investments, compares fund performance, or monitors personal trading performance. It is usually quoted as the percentage between the peak and the subsequent trough. If a trading account has $10,000 in it, and the funds drop to $9,000 before moving back above $10,000, then the trading account witnessed a 10% drawdown.

For the present case, consider the drawdown as the maximum drop on the ROI at any moment during the investment period.

2.5. Technical Indicators

A technical indicator is a mathematical pattern derived from historical data used by technical traders or investors to predict future price trends and make trading decisions. It uses a mathematical formula to derive a series of data points from past price, volume, and open interest data.

In this case, for simplicity, only one technical indicator, the relative strength index (RSI), will be considered to make trading decisions.

2.5.1. Relative Strength Index (RSI)

The RSI, a momentum oscillator developed by J. Welles Wilder, measures the speed and change of price movements. The RSI moves up and down (oscillates) between zero and 100. When the RSI is above 70, it generally indicates overbought conditions; when the RSI is below 30, it indicates oversold conditions. In this problem, the goal is to obtain the optimal values limits for the overbought and oversold conditions, as well as, the optimal RSI period, for both long and short positions.

Notice, that the RSI also generates trading signals via divergences, failure swings, and centerline crossovers.

Next the calculation of the RSI values is described.

$$RSI = 100 - \frac{100}{1+RS} \qquad \text{where} \qquad RS = \frac{AverageGain_{overNDays}}{AverageLoss_{overNDays}}$$

There are different alternatives to compute the AverageGain and AverageLoss terms, such as, Simple Moving Average, Exponential Moving Average, Wilder's Smoothing Method, etc. For simplicity, in this case, consider that both AverageGain and AverageLoss are obtained using the Simple Moving Average. Next, the calculation of the RSI for a 7-day period is illustrated.

The calculations for average gain and average loss are simple 7-period averages:

Average Gain = Sum of Gains over the past 7 periods / 7.

Average Loss = Sum of Losses over the past 7 periods / 7

Fig. 2.4, illustrates the RSI calculations considering a 7-day period, taking into account the close values from AAPL data. Assuming the overbought and oversold values as 30 and 70, respectively, the green dots represent the entry and close of a long position. Similarly, the red dots represent the entry and close of a short position.

In this case the ROI, as described in 2.3, would be:

$$ROI_{Total} = \left(ROI_{Long} + ROI_{Short}\right)/2 = (4.71+1.94)/2 = 3.32\%$$

The value of 3.32% within a range of 1 month is quite good, but, is this optimal? Is it sustainable? how risky can this simple strategy be? As you will see the applied Computational Intelligence techniques will contribute to answer these questions.

Date	Close		Gain	Loss	Average Gain	Average Loss	RS	RSI
01/08/2023	195.61							
02/08/2023	192.58		0.00	3.03				
03/08/2023	191.17		0.00	1.41				
04/08/2023	181.99		0.00	9.18				
07/08/2023	178.85		0.00	3.14				
08/08/2023	179.8		0.95	0.00				
09/08/2023	178.19		0.00	1.61				
10/08/2023	177.97		0.00	0.22	0.14	2.66	0.05	4.86
11/08/2023	177.79		0.00	0.18	0.14	2.25	0.06	5.69
14/08/2023	179.46		1.67	0.00	0.37	2.05	0.18	15.46
15/08/2023	177.45		0.00	2.01	0.37	1.02	0.37	26.79
16/08/2023	176.57		0.00	0.88	0.37	0.70	0.53	34.84
17/08/2023	174		0.00	2.57	0.24	1.07	0.22	18.27
18/08/2023	174.49		0.49	0.00	0.31	0.84	0.37	26.93
21/08/2023	175.84		1.35	0.00	0.50	0.81	0.62	38.36
22/08/2023	177.23		1.39	0.00	0.70	0.78	0.90	47.30
23/08/2023	181.12		3.89	0.00	1.02	0.78	1.30	56.60
24/08/2023	176.38		0.00	4.74	1.02	1.17	0.87	46.51
25/08/2023	178.61		2.23	0.00	1.34	1.04	1.28	56.12
28/08/2023	180.19		1.58	0.00	1.56	0.68	2.31	69.75
29/08/2023	184.12		3.93	0.00	2.05	0.68	3.03	75.20
30/08/2023	187.65		3.53	0.00	2.36	0.68	3.49	77.74
31/08/2023	187.87		0.22	0.00	2.20	0.68	3.24	76.44
01/09/2023	189.46		1.59	0.00	1.87	0.68	2.76	73.40
05/09/2023	189.7		0.24	0.00	1.90	0.00	INF	100.00
06/09/2023	182.91		0.00	6.79	1.58	0.97	1.63	62.02
07/09/2023	177.56		0.00	5.35	1.36	1.73	0.78	43.93
08/09/2023	178.18		0.62	0.00	0.89	1.73	0.51	33.81
11/09/2023	179.36		1.18	0.00	0.55	1.73	0.32	24.08
12/09/2023	176.3		0.00	3.06	0.52	2.17	0.24	19.28
13/09/2023	174.21		0.00	2.09	0.29	2.47	0.12	10.55
14/09/2023	175.74		1.53	0.00	0.48	2.47	0.19	16.15
15/09/2023	175.01		0.00	0.73	0.48	1.60	0.30	22.87

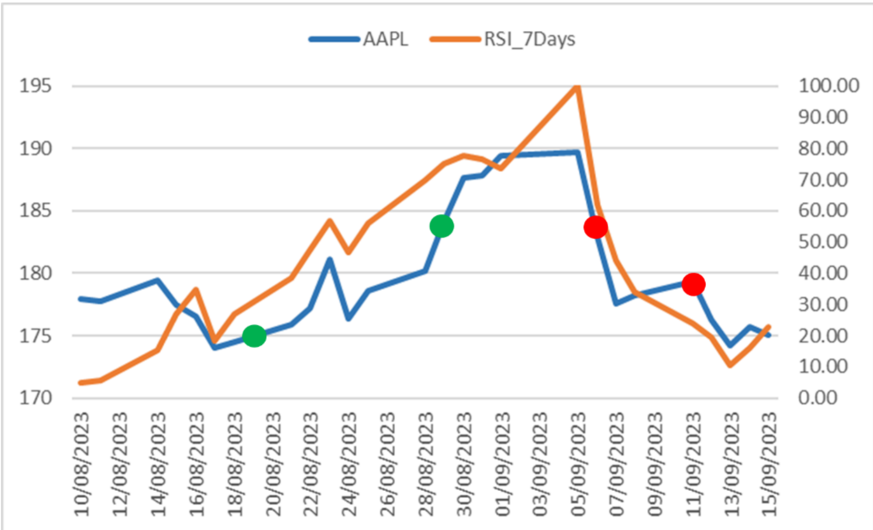

Fig. 2.4: RSI calculation and plot against AAPL data.

3. The Problem of Investing in the Stock Market (Implementation)

3.1. Data Set

The stock market data to be considered in this project is composed by 10 data series, each one corresponding to a stock and including data from January 2006 to December 2022. Each data series is described as a CSV file named as described in table 3.1. The file structure is illustrated in table 3.2, for the case of the Apple data series, in this project just columns A (Date) and E (Close) will be considered. Column A indicates the date and column E indicates the "Close", which means the last value of the stock on that date.

Table 3.1: Stock file names and data series size

Filename	Stock Name	#Days	Filename	Stock Name	#Days
AAL.csv	American Airlines	4523	GOOG.csv	Alphabet	4802
AAPL.csv	Apple	10780	IBM.csv	International Business Machines	15533
AMZN.csv	Amazon.com	6628	INTC.csv	Intel	10968
BAC.csv	Bank of America	12754	NVDA.csv	NVIDIA	6203
F.csv	Ford Motor Company	12934	XOM.csv	Exxon Mobil	15533

Table 3.2: Apple CSV file sample

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	12/12/1980	0.128348	0.128906	0.128348	0.128348	0.09945	4.69E+08
3	15/12/1980	0.12221	0.12221	0.121652	0.121652	0.094261	1.76E+08
10780	14/09/2023	174	176.1	173.58	175.74	175.74	60895800
10781	15/09/2023	176.48	176.5	173.82	175.01	175.01	1.09E+08

3.2. SOO Applied to Maximize Return on Investment (ROI) using Technical Indicators

The single objective optimization applied to maximize return on investment (ROI) using technical indicators **must consider**:

- Maximize ROI as the objective function.
- the Relative Strength Index (RSI) as technical indicator, where the parameters to be optimized are:
  - lower band value, to open a long position (values from 0 to 100, multiples of 5)
  - upper band value, to close a long position (values from 0 to 100, multiples of 5)
  - the RSI period (nº of days), to apply for long positions (7, 14 or 21 days)
  - lower band value, to close a short position (values from 0 to 100, multiples of 5)
  - upper band value, to open a short position (values from 0 to 100, multiples of 5)
  - the RSI period (nº of days), to apply for short positions (7, 14 or 21 days)
- 10000 evaluations as the stop criteria
- the complete data series values from Jan 2020 until Dec 2022. (3 years)

This means that the solution space has 1440000 candidates! Justify!

First, start by defining how to represent the candidate solutions and identify which would be the evolutionary operators to consider. Afterwards, select an evolutionary approach and use either functions from an existing library or implement the EA from scratch (both solutions have their pros and cons … multi-objective problem 😉)

Complete a table indicating the EA approach considered and the chosen hyper parameters for that approach.

Table 3.3: SO Approach and Parameters

EA Approach:	*EAname*	------------
Hyper Parameter 1	*HP1name*	*HP1 value and/or method*
Hyper Parameter 2	*HP2name*	*HP2 value and/or method*
…	…	…
Hyper Parameter n	*HPnname*	*HPn value and/or method*

Complete the following table based on the execution of 30 runs (with different random seed) for each stock.

Table 3.4: ROI (Max, Min, Mean and STD) over 30 runs

Stocks	ROI (2020-2022)				Stocks	ROI (2020-2022)			
	Max	Min	Mean	STD		Max	Min	Mean	STD
AAL.csv					GOOG.csv				
AAPL.csv					IBM.csv				
AMZN.csv					INTC.csv				
BAC.csv					NVDA.csv				
F.csv					XOM.csv				

Generate a boxplot graph with the results obtained each of above runs (suggestion use a normalization so that the boxplot for different stocks can be compared). Comment results.

Create 6 histograms, one for each of the optimization variables with all the achieved results on the above experiments. Each histogram should have 300 values (30 (runs) x 10 (stocks)), as in the boxplot analysis use normalization when generating. Comment results.

3.3. Train & Test Scheme

The ROI results obtained on the previous section can be optimal for the presented data series but are not realistic once it considers future values to optimize the RSI parameters. The investor has no information on future values, e.g., when deciding to invest in 2021 he/she cannot rely on optimal parameters obtained considering values until 2022. So, a more realistic approach is to define a **training period** to obtain the optimal parameters and then apply those parameters in a **test period**.

In order to obtain more realistic values, consider a scheme where 9 years are used for training and 3 year for test, i.e. train period from Jan 2011 until Dec 2019 and test period from Jan 2020 until Dec 2022.

Considering the same conditions in the previous section, **complete the following table** based on the execution of 30 runs (with different random seed) for each stock and for **training period**.

Table 3.5: ROI in the Train Period (Max, Min, Mean and STD) over 30 runs

Stocks	ROI (2011-2019) – Train Period				Stocks	ROI (2011-2019) – Train Period			
	Max	Min	Mean	STD		Max	Min	Mean	STD
AAL.csv					GOOG.csv				
AAPL.csv					IBM.csv				
AMZN.csv					INTC.csv				
BAC.csv					NVDA.csv				
F.csv					XOM.csv				

For each of the above 30 runs, for each stock, apply the obtained values for the optimization parameters and apply during the test period, **complete the following table**.

Table 3.6: ROI in the Test Period (Max, Min, Mean and STD) over 30 runs

Stocks	ROI (2020-2022) – Test Period				Stocks	ROI (2020-2022) – Test Period			
	Max	Min	Mean	STD		Max	Min	Mean	STD
AAL.csv					GOOG.csv				
AAPL.csv					IBM.csv				
AMZN.csv					INTC.csv				
BAC.csv					NVDA.csv				
F.csv					XOM.csv				

Compare and comment the results on tables 3.4 and 3.6.

3.4. MOO Applied to Maximize Return on Investment (ROI) and Minimize Drawdown using Technical Indicators

The multi objective optimization applied to maximize return on investment (ROI) and minimize the Drawdown using technical indicators **must consider**:

- **Maximize ROI** as one objective function and **minimize drawdown (DD)** as other objective function.
- the Relative Strength Index (RSI) as technical indicator, where the parameters to be optimized are:
  - lower band value, to open a long position (values from 0 to 100, multiples of 5)
  - upper band value, to close a long position (values from 0 to 100, multiples of 5)
  - the RSI period (nº of days), to apply for long positions (7, 14 or 21 days)
  - lower band value, to close a short position (values from 0 to 100, multiples of 5)
  - upper band value, to open a short position (values from 0 to 100, multiples of 5)
  - the RSI period (nº of days), to apply for short positions (7, 14 or 21 days)
- 10000 evaluations as the stop criteria
- the complete data series values from Jan 2020 until Dec 2022. (10 years)

This means that the solution space has 1440000 candidates!

Complete the following table for each stock, considering a population size of 64.

Table 2.7: ROI and DD for the MO approach over 30 runs

Stocks	Max ROI		Min Drawdown		Stocks	Max ROI		Min Drawdown	
	ROI	DD	ROI	DD		ROI	DD	ROI	DD
AAL.csv					GOOG.csv				
AAPL.csv					IBM.csv				
AMZN.csv					INTC.csv				
BAC.csv					NVDA.csv				
F.csv					XOM.csv				

Generate the Pareto Front graphs for each of the above tests and superimpose them in a single plot.

Suggest and justify a train and test scheme when using a multi-objective approach.

4. Work Submission

The work should be **submitted on the course website by October 27th, 23:59**. The submission must include a zip package with: (1) the **code (commented)** to be tested in lab environment, also, include a **README file** describing the use and the sequence of results the user will see when testing your work; (2) the **report** should include (2.1) a **brief graphical description of the chosen algorithms and problem representation for each problem** (single and multiple objective), (2.2) the **tables and graphs from sections 3.2, 3.3 and 3.4,** and (2.3) a final section with **concluding remarks**. The report **should not exceed 5 pages, including the cover page**.

5. Evaluation

The work evaluation is based on the work submitted (50%) and on the oral discussion (50%).
The **single objective part** corresponds to 3/4 of the work classification (15 over 20, 12 over 15 for section 3.2 and 3 over 15 for section 3.3)
The **multiple-objective part** corresponds to 1/4 of the work classification (5 over 20).