Project 2: Genetic Algortihm Makes Money
Due Date: 02/19/2018 at 12:00pm

1 Design

1.1 Goal

The goal of this project is to use a genetic algorithm to find a simple and good rule that could be used in Stock Market trading. The genetic algorithm is to implemented in Java.

1.2 Rules

Three basic trading rules are used in this project. The first basic rule uses the Sample Moving Average (SMA) calculation. In this rule, if the actual closing price of a share or stock is higher than the SMA over N periods of time, then a buy signal is sent; otherwise a sell signal is sent. The second basic rule locates the maximum closing price from a period of N days. If this maximum is less than the actual closing price, then a buy signal is sent; otherwise a sell signal is sent. The third basic rule uses the Exponential Moving Average (EMA). If the actual closing price of a share or stock is higher than the EMA over N periods of time, then a buy signal is sent; otherwise a sell signal is sent.

1.3 Individual in the population

Each individual represents a rule. Individuals can either be one of the basic rules defined above or a combination of the basic rules with the logical connectors: AND or OR. For example, one individual can be "SMA 50". This means that the SMA over a period of 50 days will be calculated, and if it is less than the actual closing price, then a buy signal is sent. Another example of individual can be "SMA 20 AND MAX 30", which combines the first two basic rules together.

1.4 Individual representation

To ease up the computation, individuals will be presented in a string of characters. The character 's' is the substitute for SMA, 'e' for EMA, 'm' for MAX, '&' for AND, and '|' for OR. The numerical values remain as they are in the individual representation. An Individual in the population will have a length of 14 characters. The first example of an individual, in the previous paragraph, will be shown in the following form: s050&e000&m000. Rules are evaluated from left to right; therefore, the following rule, SMA 50 AND EMA 30 AND MAX 10, represented by the individual s050&e030&m010, is evaluated as follow (SMA 50 AND EMA 50) AND MAX 10.

1.5 Crossover

The two individuals that will take part in the crossover will be selected using a roulette selection. The crossover point will be chosen randomly from the interval (0, 13). The two offspring from the crossover will both be evaluated and then added to the initial population. The crossover rate (probability) will be fixed at 0.8.

1.6 Mutation

Mutation will be performed by changing a character in the string that represents an individual with another appropriate character. An 's' for example can be replaced by an 'e' or an 'm' and vice versa. Any digit can be replaced by any digit from 0 to 9. The '&' character can be replaced by an '|' character and vice versa. Each character in the string will be changeable under a defined mutation rate. The mutation rate(probability) will be fixed at 0.001.

1.7 Fitness evaluation

Individuals will be evaluated based on the performance of the rule that they represent. Rules will be applied to 5 different historical prices from 5 different companies (Ford, Apple, NATI, Nike, and Google). A starting amount of $20,000 will be used as a startup deposit for each of the 5 trade groups. The gains or profits obtained after selling shares will be taken out of the account and will be kept in a gain account. If the account happens to go under $20,000 and if there are enough money in the gain account to get the account back to $20,000, then some money will be moved out of the gain account to get the account back to $20,000. A starting and ending date will be set for the trades. When the ending date is reached and if there are still some shares left in the share account, then all of these shares will be sold using the ending date's closing price. Rules that did not buy or sell shares at all during the entire trading period will be penalized by reducing the money in their account by half. This penalization will be created to make sure that the genetic algorithm will not converge to rules that tend to not participate in the market. A transaction fee of $7 will be applied to a buy and a sell action. The final balances, in the 5 different accounts for the 5 different trade groups, will be summed together. This sum will be considered as the fitness of the rule or the individual.

1.8 Halting condition

The genetic algorithm will have 20 individuals int its initial population, and then it will produce 200 new generations from the initial population before it will stop and output the best fit individual.

2 Experiment

2.1 Rule Testing

The rule outputted by the genetic algorithm will be tested following the way how its fitness was obtained with small changes in the settings. The rule will be applied to 30 historical prices that are all different from each other and different from the 5 historical prices used in the genetic algorithm. The rule will be given a starting account of $100,000 for each historical price instead of $20000. The gain that the rule will make after trading under one historical price will be recorded on a table. Like in the genetic algorithm, the rule will be forced to sell at the end of the trading period if some shares are left unsold. Unlike in the genetic algorithm, rules will not be penalized if the do not buy or sell during the trading period.

3 Deliverable

Along with submitting your code, write a paper that summarizes what you have discovered from the experiment. Use tables and graphs as necessary. The structure of your paper should follow the following outline:

- Introduction
- The process/project that you are implementing
- Experiment and Results
- Conclusion and Discussion

4 Appendix

$$\text{EMA}_{\text{today}} = \frac{p_1 + (1-\alpha)p_2 + (1-\alpha)^2 p_3 + (1-\alpha)^3 p_4 + \cdots}{1 + (1-\alpha) + (1-\alpha)^2 + (1-\alpha)^3 + \cdots}$$

$$\alpha = 2/(N+1)$$

$$SMA = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$

Go here to download the historical data for companies, `https://finance.yahoo.com/lookup?s=YHOO`. In the Genetic Algorithm, the minimum time period should be 1 year and the maximum time period should be 3 years. In the experiment (i.e rule testing) the time period should be exactly 3 years.