Applications of Machine Learning to Stock Market Predictions

Anson Wong, Gudbrand Duff Morris and Juan Garcia

February 28, 2017

Introduction

We propose to undertake an exploratory project on the potential impact of machine learning techniques to stock market prediction and analysis.

The idea of using algorithms to predict the price of assets is as old as the stock exchange itself, and the existing body of work on the subject is huge. Therefore, **the first part of our project will be an extensive literature review**. We will read and share interesting papers, using our findings to guide the development of the project scope and details.

Each one of the participants in this project have a possible idea in what direction this project could go:

Gudbrand's driving question: With the dawn of the Era of Big Data upon us, I seek to discover if and how a "change in magnitude will produce a change in kind".

Anson's asks: With the lack of insider information for amateur traders and the rise of algorithmic traders in the current stock market, is trading a lost cause for you and I? If not, then could an amateur machine learnist take a stab at trading and forecast price action with high success probability? I believe this is possible to earn large gains in niché markets using a healthy amount of machine learning and understanding of human psychology. We plan to explore reinforcement learning and feature selection to achieve such a goal.

Juan's Idea: A machine can perform better than a human in a wide range of situations. Forecasting and detecting patterns in the time series in the stock market

could be one of these situations. My idea is that with the techniques and ideas we are learning, we can teach a machine how to recognize what a good investment is. The criteria to define 'good' comes from our knowledge from stock markets after doing some literature review. For the moment my first though in that direction is to apply the classification techniques we are learning, in order to select the portfolio with less variance. In this case the concept of 'good' investment is 'safe' investment.

Our literature research so far is summarized below

Literature Review

Papers:

- http://www-stat.wharton.upenn.edu/ steele/Courses/434/434Context/EfficientMarket/AndyLoJPM2004.pdf

- http://www-stat.wharton.upenn.edu/ steele/Courses/434/434Context/EfficientMarket/Granger-stockmarket.pdf

Interesting/useful links

- This guy is pretty great: `http://www-stat.wharton.upenn.edu/~steele/`

- `https://www.udacity.com/course/machine-learning-for-trading--ud501`

- `https://www.quantopian.com/posts/simple-machine-learning-example`

- `http://www.qminitiative.org/UserFiles/files/S_Cl%C3%A9men%C3%A7on_ML.pdf`

- `http://www-stat.wharton.upenn.edu/~steele/Courses/9xx/Resources/MLFinancialApplications/MLFinance.html`

- `https://www.quora.com/How-do-financial-companies-use-machine-learning`

- `http://techemergence.com/machine-learning-in-finance-applications/`

- Our GoogleDrive: `https://drive.google.com/drive/folders/0B5_P6FZEZzQxOEN6ZTU1Q1czQTQ?usp=sharing`

- R package for financial analysis: `https://cran.r-project.org/web/views/Finance.html`

Applications/Implementations/Ideas

As the famous Oscar Wilde quote goes; "A cynic is a man who knows the price of everything but the value of nothing". In this sense, a ML model can be regarded as the ultimate cynic. The cynic has both advantages and disadvantages compared with the expert investor. With this in mind we have a set of potential topics/ideas to start with:

Predictability How hard is the problem? Almost a philosophical question.

Clustering What sort of stocks/bonds/instruments are there? Which models work for different categories?

Feature Selection Which features are important when it comes to prediction? We see many examples of seemingly godly predictions made based on the most complex features. Is this sort of 'thinking' possible for machines? Take for example the fictional Silicon Valley investor Peter Gregory, who reasons from the popularity of Burger King to the impending surge in periodic cicadas to Indonesian sesame futures. Or hedge fund manager Micheal Burry, one of the few big shot investors to recognize the subprime mortgage crisis long before it happened, just by analyzing the right data.

Prediction techniques How can we predict the behaviour of a stochastic PDE (Black Scholes')?

Momentum Investing Gradient step based on today's winners. A potential winning strategy.

Page Rank Page rank is particularly useful for ranking connections (whether it be for search results, biological systems, social media influence). Although it might be a stretch, are there ways to use Page Rank to shed light on how the stock market works?

Q-learning A method of reinforcement learning which sets up a reward system, and uses a mix of exploration and exploitation to achieve optimization of a

goal. Planning to apply this to create a trading monkey.

Financial Background The introduction should get the reader up to date on the subject. Efficient markets, price, information, volatility, portfolio, ...

Portfolio Selection Real-time optimization++

Finally we consider some technical aspects regarding data sets and software.

Other

- Datasets: There are plenty of sites were time series for the stock market can be accesed, such as, Finviz.com, Google Finance, Yahoo Finance, MarketWatch, to name a few.

- Programing Language and Packages: Any high level programing can be use to do the data analysis. At this point in the project there is some preference for R and the financial analysis packages such as Empirical Finance. This is since there is familiarity from the members of this research group with this language.