Machine Learning for Stock Price Prediction

Juan Garcia
99999999

Gudbrand Tandberg
83628164

Anson Wong
99999999

Abstract

In this paper, we investigate the performance of machine learning models for the task of stock price prediction. After reviewing the basics of stock market prediction, we present an ensemble model that utilizes a variety of ML-techniques to forecast the future price of a given stock.

1 Introduction

At the heart of classical market theory lies "The Efficient Market Hypothesis", which states that markets are *efficient*:

$$H_{01} : \text{A market is efficient with respect to information set } I_t \text{ if it is impossible}$$
$$\text{to make economic profits by trading on the basis of this information set.}$$

A slightly stronger formulation of this hypothesis is

$$H_{02} : \text{Stock prices are a martingale}$$
$$\text{i.e. } E[P_{t+1} \,|\, I_t] = P_t.$$

The mere existence of profitable investment funds provide enough counter-evidence to this hypothesis to suggest the following partial refutation of the EMH:

$$H_1 : \text{there exist at least } some\ times \text{ where at least } some\ part \text{ of the market is } inefficient".$$

Working from this hypothesis, market investors attempt to predict markets using a combination of three broad categories of prediction methodologies: fundamental analysis, technical analysis, and data mining technologies. Fundamental analysts are concerned with the company that underlies the stock itself. Technical analysts are not concerned with any of the company's fundamentals. Instead, they seek to determine the future price of a stock based solely on the (potential) trends of the past price (a form of time series analysis). With the advent of the digital computer, stock market prediction has since moved into the technological realm, where for example internet-based data sources and machine learning algorithms are used for predicting the future.

The stock market is essentially dynamic, non-linear, complicated, nonparametric, and chaotic in nature. The time series are multi-stationary, noisy, random, and has frequent structural breaks. In addition, stock market?s movements are affected by many macro-economical factors such as political events, firms' policies, general economic conditions, commodity price index, bank rate, bank exchange rate, investors' expectations, institutional investors? choices, movements of other stock market, psychology of investors, etc. It is our belief that the only reasonable way of overcoming all these difficulties is by harnessing the power of big data and modern machine learning algorithms and

architectures. Perhaps even it is possible to devise models of stock prediction that will allow fully automated investment funds to operate with a higher success than funds reliant on human "expert knowledge".

You should explicitly state your contribution in the introduction of the paper,

2 Related Work

This is a citation Vapnik [1999].

3 The Data

3.1 Data Cleaning

3.2 Feature Engineering

4 Our Model

4.1 SVM Regression

4.2 Gaussian Processes

4.3 Markov Model

4.4 ...

4.5 The Ensemble Learner

5 Results

6 Discussion

State the main conclusions that are obtained from this course project. List at least one strength and one weakness of your contribution. Briefly state what you would do with more time

References

Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.