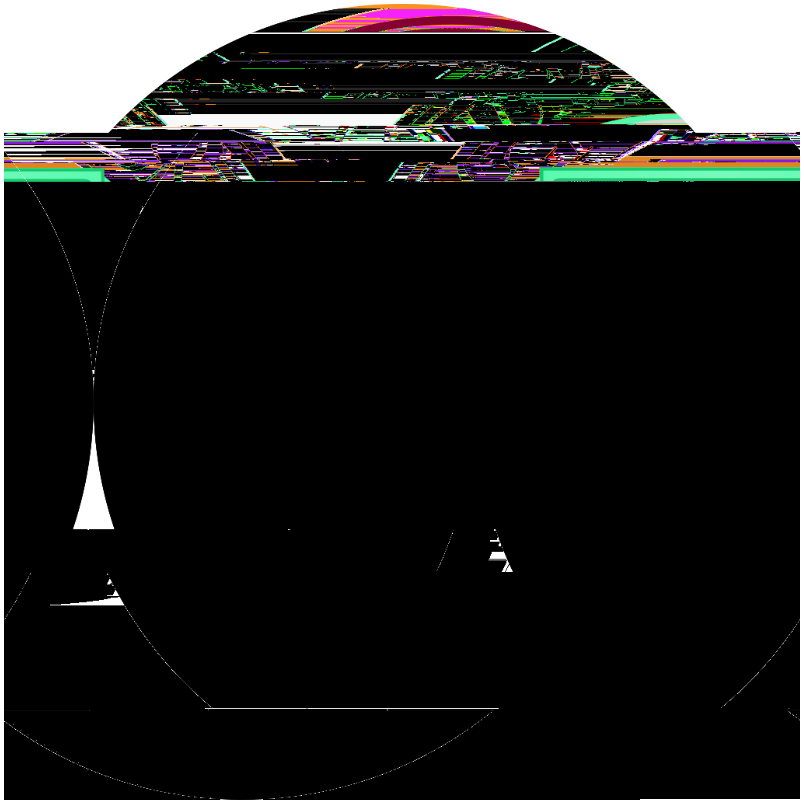
# A Machine Learning Approach to Automated Trading

Boston College Computer Science Senior Thesis



Ning Lu

Advisor: Sergio Alvarez

May 9, 2016

# Abstract

The Financial Market is a complex and dynamical

# Contents

# 1. Introduction

As of 2012, a report from Mo

# 2. Automated Trading

## 2.1 Overview and Terms

The Stock Market is a marketplace where shares of public companies are traded. A company becomes public when it makes an Initial Public Offering, or commonly known as IPO, which means that investors worldwide are able to buy and trade shares of stock in the company. These shares represent part ownership in the company, and their prices represent what investors believe a orders everyday to seek intraday profit. For automated trading systems, because of their systematic nature, the trading frequency needs to be specified by their developers before even designing the them. Trading frequency has very wide range: from once a lifetime, which means buy and hold forever, to once every nanosecond for High-Frequency Trading algorithms. As a result, mo system minimizes emotions throughout the trading process, ensures that discipline is maintained through volatile market conditions, and allow us to achieve consistency.

The other advantage of algorithmic trading over human traders is the ability to backtest the stm rR

## 2.3 Past Approaches

In the domain of using machine learning techniques to forecast stock market movements, Financial Time Series Forecasting with Machine Learning Techniques: A Survey [17] provides a cohesive overview of recent developments: Out of 46 publications this survey covers, 21 of them use Arti

# 3. Machine Learning Methods

## 3.1 Brief Introduction

Alan Turing, one of the greatest computer scientists in human history and father of artificial intelligence, in his paper ñComputing Machinery and Intelligenceò[33] asks this famous question: ñCan machines think?ò Until today, we as human still donõt know the answer to this question, but variati

## 3.2 Basic Models & Variations

### 3.2.1 Logistic Regression

Logistic regression was developed by statistician David Cox in his 1958 paper [8]. It is a special case of generalized linear model, and thus analogous to linear regression. The idea is to estimate the conditional probability of a binary response based on one or more predictor variables by using the cumulative logistic distribution.

If we have a binary output variable Y, and we want to model the conditional probability P(Y=1|X=x) as a function of a vector x. Notice that log p(x) is a linear function of x, which has an unbounded range, but a easy modification, the logistic transformation log p/(1-p) is bounded. Therefore, the logistic regression model is

$$log \frac{p(x)}{1-p(x)} = \beta_0 + x \cdot \beta \qquad (1)$$

And solving for p, gives a form of the sigmoid function.

$$p(y|x) = \frac{1}{1+e^{-(\beta_0 + x \cdot \beta)}} \qquad (2)$$

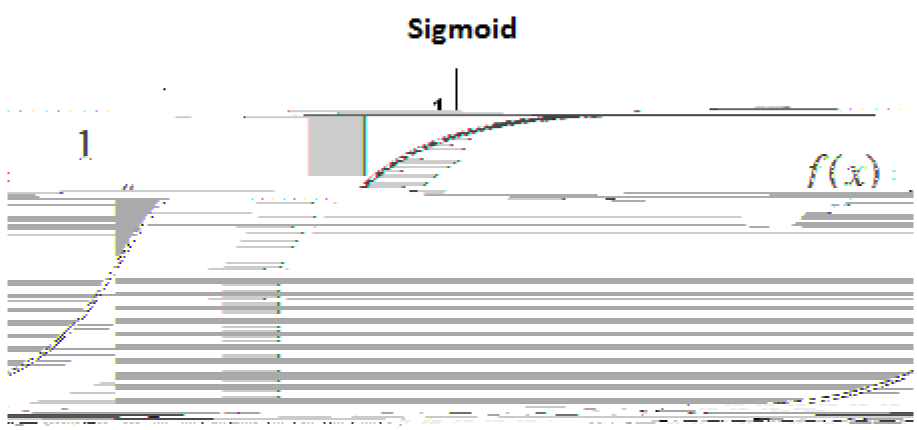



Figure 1: The logistic, or sigmoid, function.

Two variations of logistic regression are implemented for the purpose of this paper. Logistic regression with a ridge penalty is used for the individual approach, and the lasso logistic regression is used for the sector approach. The applications will be discussed in greater detail later.

Ridge regression minimizes

$$\sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 + \lambda \sum_j \beta_j{}^2 \tag{3}$$

Where N is the number of observations, $y_i$ is the response at observation i, $x_i$ is the vector of independent variables at observation i, $_j$ are the coefficients of the original logistic regression, and lambda is the positive regularization parameter, also known as the ridge penalty. The second term shrinks the coefficients to prevent any one of the them being too large and cause overfitting.

Similarly, the lasso regularization [32] minimizes

$$\sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 + \lambda \sum_j |\beta_j| \tag{4}$$

Instead of having the square term, the absolute value term is known as a $L_1$-norm penalty. By increasing lambda, the lasso penalty also force the coefficients to shrink, but in this case they tend to be truncated at 0. Therefore, the lasso regularization can work as a feature selection process.

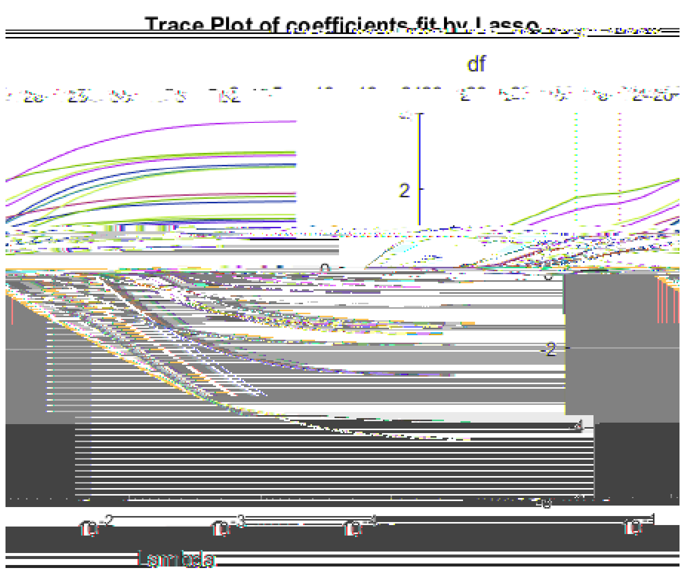



Figure 2: Example of the lasso logistic regression selecting meaningful coefficients

### 3.2.2 Decision Tree

A decision tree is a hierarchical data structure implementing the divide-and-conquer strategy, whereby the local region is identified in a sequence of recursive splits in a smaller number of steps [1]. Tl

However, a deep tree with many leaves still tends to overfit and therefore its test accuracy is often far less tha

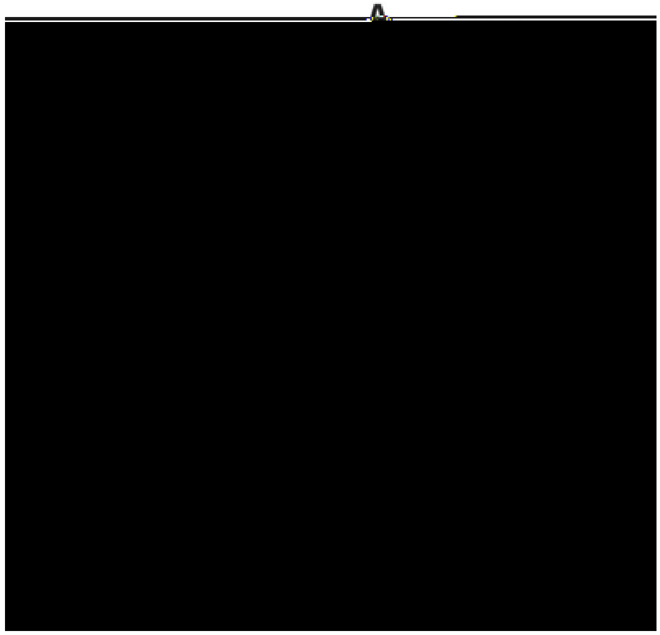


Figure 4: An example of ROC curve and its optimal threshold

ROC curve in general is used to compare performances across different classifiers, by comparing each classifier's AUC (area under curve) of its ROC curve. But in this case, we use the ROC curve to fi on the length of the margin for every observation that is on the wrong side of its class boundary when classes are inseparable.

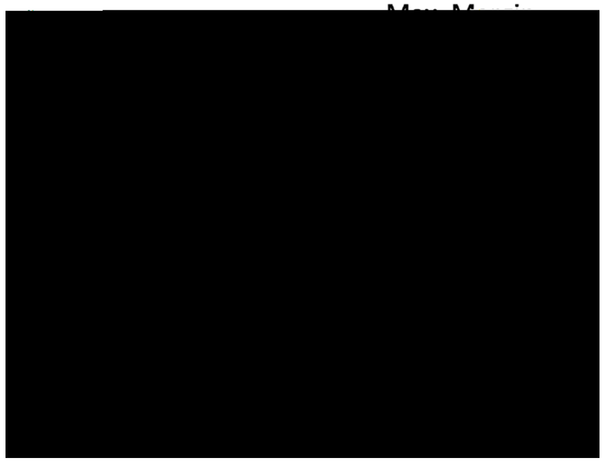


Figure 5: an example of a Support Vector Machine in a separable case

In addition to performing l

# 4. Prediction Model: Individual Approach

## 4.1 Methodology

The first method to develop a stock price prediction model is what we called an individual approach, which means that the model uses historical performance of a stock itself to predict its future price movements. The idea is that, looking at how the stock historically moves in N-day win`

## 4.2 Data

The data we use  is web scraped from Yahoo Finance, which contains adjusted closing price for all the stocks in S&P 500 from Jan 2,2014 to Feb 1, 2016. We did not go further back because prices more than 2 years old is likely to be irrelevant. Adjusted closing price is the daily closing price, which is the last trading price before market closes at 4 pm, adj

## 4.3 Results

### 4.3.1 Evaluation Metric

There are two ways that we use in this thesis to evaluate the performance

| | Logistic Regression | | | Ridge Logistic Regression | | |
|---|---|---|---|---|---|---|
| Threshold | TP | TN | TR | TP | TN | TR |
| 0.5 | 0.5002(0.0552) | 0.5118(0.1051) | 0.5067(0.0421) | 0.5032(0.0590) | 0.5086(0.0560) | 0.5050(0.0362) |
| 0.55 | 0.5192(0.1765) | 0.5006(0.2767) | 0.5161(0.1447) | 0.5680(0.3346) | 0.4998(0.3501) | 0.5289(0.3141) |
| | | | | | | |

# 5. Prediction Model: Sector Approach

## 5.1 Methodology

The idea of the sector approach is to allow our model to look at a bigger pic

From these six models, four basic and two ensemble, we will pick the top performing ones to backtest on real out-of-sample data in the next section.

## 5.2 Da

## 5.3 Results

### 5.3.1 True Rates

Similar to the approach in the previous section, we split the dataset into 80 percent training and 20 percent testing. Since training error means nothing when implementing the

| | True Positive | True Negative | True Rate |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

as follows: if the original Naive Bayes predicts up, the final model takes this prediction, but if the original model predicts down, the final model only makes the same prediction if the probability of predicting a negative return given by the original model exceeds the optimal threshold from the ROC curve.

| | True Positive | True Negative | True Rate |
|---|---|---|---|
| Utility | 0.5949 (0.0623) | 0.5023 (0.0623) | 0.5513 (0.0456) |
| Energy | 0.4797 (0.0467) | 0.5884 (0.0671) | 0.5210 (0.0335) |
| Information Technology | 0.5115 (0.0523) | 0.5057 (0.0595) | 0.5096 (0.0346) |

Table 8: Average true rates of Naive Bayes with ROC analysis on different sectors

The true negative rates improve slightly using the two-step rule. Although we believe this impr[redacted]ment may not be significant, it is good to see no signs of overfitting.

[redacted]ext, we look at the true rates from running the two ensemble methods. Again, R

[redacted]

However, some of the true rates with p-value larger than 0.01 cannot reject the null hypothesis at 1% significance level. Also, the t-tests confirm our previous finding that SVM and Naive Bayes have better upside predictions than downside, since the true negative rates from both utility and information technology sectors fail to be different than random guessing.

| | True Positive | True Negative | True Rate |
|---|---|---|---|
| Utility | 0 | 0 | 0 |
| Energy | 0 | 0 | 0 |
| Information Technology | 0 | 0 | 0 |

Table 17: Lilliefors test results of Random Subset using Naive Bayes with ROC analysis

| | True Positive | True Negative | True Rate |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

# 6[redacted]

## 6.1 Str[redacted]

[redacted]cond [redacted] that u[redacted]achine lea[redacted] not e[redacted] simply buy ev[redacted] sign[redacted]. A good trading str[redacted] pre[redacted]by understanding the ess[redacted] wh[redacted]rate trade orders.

[redacted]e models proposed in this paper a[redacted] s[redacted]sence is that we can only trust the mod[redacted] need to take this in consideration. To illustra[redacted]ward predictions for the past five days and we[redacted]ach day. The position in this stock we have at t[redacted]e model gives us a sell signal for day six. A qu[redacted] we want to sell given this sell signal. If w

The above figure demonstrates the strategy we designed specifically for the models proposed in this paper. This strategy takes in consideration the existing portfolio positions before submitting a trade order. Let's say the model gives a buy signal, which means that it suggest the stock price is going up tomorrow. If we already hold a long position, which cumulates from previous buy signals, or have no position at all, it makes sense to buy one more unit or enter a long position. But if we are shorting the stock, we know that the short position will suffer lost tomorrow when price goes up, so we need to exist all short position to avoid this lost, and, we want to go aggressive and even start entering a new long position to profit from the price increase. This aggressiveness depends highly on our model's accuracy and as a result, the returns generated by this strategy are meant to have higher volatilities.

In Section 4 and 5, we train the Machine Learning prediction models on the first 70 percent of the data and calculate the true rates on the rest 30 percent. In real world situation, however, new data point comes in every day so we can also take advantage of this fact. Instead of building just one training data set, we build it dynamically, that is, our trading system update the train set everyday by adding the new entry.

Moreover, for each sector, we pick the top 10 performers from the validation period to trade on. The table below shows the tickers for these best performing stocks. Only using the top 10 stocks may cause overfitting thus hurt the out-sample results. In real sit

## 6.2 Market Simulator

We develop a market simulator to provide an environment where the trading strategies can be tested. The simulator uses real-work stock prices, and inherits a simplified version of real markd

## 6.3 Out-of-Sample Live Simulation

### 6.3.1 Sharpe Ratio

In order to measure a strategyʼs performance, looking at just the historical returns the strategy generates is not enough. We need a performance measure that considers the consistency of the returns generated by a strategy. In William F. Sharpeʼs 1994 original paper [28], He introduced the Sharpe Ratio which is a measure for calculating risk-adjusted return. The Sharpe ratio equals mean of portfolio return minus risk-free rate and divided by standard deviation of portfolio return.

The Sharpe Ratio is the average return earned in exc

### 6.3.2 Individual Approach

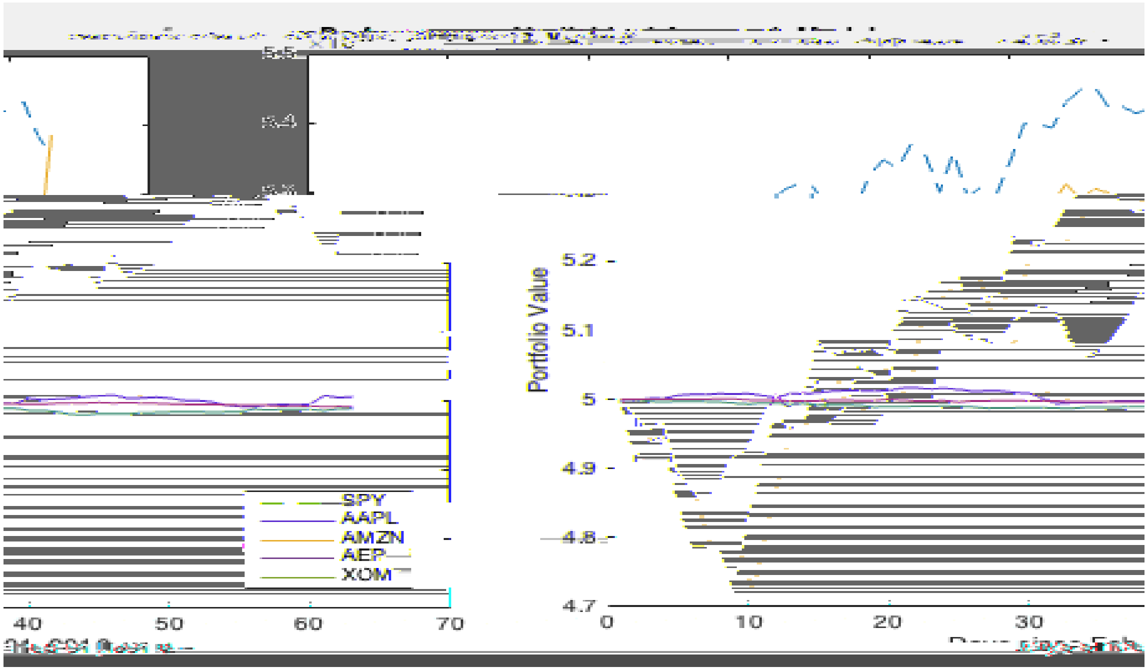



Figure 7: Three months out-of-sample simulation results for individual approach models

| | SPY | AAPL | AMZN | XOM | AEP |
|---|---|---|---|---|---|
| Sharpe | 2.0738 | 0.5343 | 2.8183 | -2.8736 | -2.0579 |

oi

Table 20: Sharpe ratio of ridge logistic regression trading on different stocks

Other Amazon, systems that trades on Apple, XOM and AEP all underperform the Market in this ou

Here, “Naive Bayes” refers to a trading system which use Naive Bayes with ROC optimization as the underlying model, and similarly with “Random Subset”, which refers to a system which employs the Random Subset technique with ROC-optimized Naive Bayes as the underlying model.

Naive Bayes on utility, energy and information technology sectors all achieve positive returns during the testing period. In particular Naive Bayes trading on the top 10 stocks in the energy sector has a promising 2.14 sharpe ratio, though the other two do not have sharpe ratios above 1. The Random Subse

# 7. Conclusion

In this th

to help us decide whether the backtest is predictive and the automated trading system will continue to be predictive.

## Acknowledgment

First of all, I want to sincerely thank my parents. Without their unconditional support, I would not have the opportunity to receive high quality education in the United States and complete a thesis at the amazing Boston College. Thank you Professor Alvarez, my thesis advisor, and the entire Computer Science department for providing me resources and giving me very helpful advices not only for this thesis work, but throughout my four years as a Computer Science major. Thank you Professor Lowrie for being an amazing academic advisor and congratulation on your retirement. I want to also thank Professor Law from the economic department, his new course Intro to Computational Investing really gave me a good start on this project, and I borrowed his idea of building a market simulator to build mine. Yang Zhou, my partner in this class has also contributed to the development of the logistic regression in this paper. Thank you Boyao Sun for always counter arguing me and helped me with the idea of the random subset method. Thank you Cydgg, Emma and Congcong for always reminding me to be a kind person. Also, thank you Professor Sadka from CSOM who also gave me advice on this thesis, and I want to thank Carnegie Mellon University for giving me the opportunity to lean more about the field of Quantitative Finance in the MSCF program. Thank you Alan Turing, and everyone who has contributed to the field of Machine Learning and continue to push the advancement of modern technology. Last and specially, I want to thank Summer Li for your immense support and understanding. I sacrificed a lot of time that belongs to you for this thesis, and I want to repay you for the rest of my life.

[28] Sharpe, William F. "The sharpe ratio." The journal of portfolio management 21.1 (1994): 49-58.

[29] Sharpe, WF. "The Sharpe Ratio - Stanford University." 2015. https://web.stanford.edu/~wfsharpe/art/sr/sr.htm

[30] Simon, Phil. *Too Big to Ignore: The Business Case for Big Data*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.

[31] Stickel, Scott E. "Predicting individual analyst earnings forecasts." Journal of Accounting Research (1990): 409-417.

[32] Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society. Series B (Methodological) (1996).