A machine learning based stock trading framework using technical and economic analysis

Smarth Behl (smarth), Kiran Tondehal (kirantl), Naveed Zaman (naveedz)

Stanford University

Introduction

Investors in the stock market can maximize their profits by buying or selling their investments if they can determine when to enter and exit the positions. Professional traders typically use fundamental and/or technical analysis to analyze stocks in making investment decisions.

Compared to existing work, this project analyzes the stocks trading decisions utilizing the technical behavior of the trading patterns within the context of the fluctuating economic and business environment.

The objective function is to maximize medium to longer term profits based on S&P500 stock market index.

Data and Features

For technical analysis, the focus has been on the most prominent indicators that can be efficiently operationalized and are intuitive in interpretation, including: Moving average convergence & divergence; Stochastic KD;Relative strength index; Larry William's R %; Daily closing volume.

For economic analysis, indicators being utilized in terms of their importance and data availability are: Gross Domestic Product; Consumer price Index; Producer Price Index; Employment Index; Fed Funds Rate.

Technical Indicators have been calculated from the downloaded daily closing prices & volume data. The closing prices & volume have been smoothened using Welles Wilder Smoothing without any look-ahead bias, and relative 15-day change is calculated to serve as the price & volume trend indicators.

Economic indicators have been extracted from the officially released historical percentage change data. In additon, Fed Funds rate have been smoothened using Welles Wilder Smoothing and relative 15-day change has been applied to capture the trends in the economic cycle.

Methodology

Several supervised learning methods were considered before deciding on the three models used for the project .

1. Neural Network

The Neural Network architecture consists of 15 hidden nodes. Sigmoid activation is used for the hidden nodes activation, softmax activation is used for the output layer.

$$z = w^T x + b$$

$$a = \frac{1}{1 + e^{-z}}$$

$$\hat{y} = \frac{\exp(z)}{\Sigma_k \exp(z)}$$

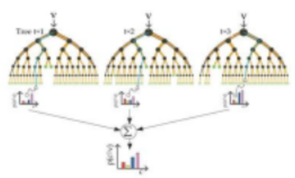
2. Softmax Logistic Regression

A 4-class logistic regression model is used.

- $p(y = i|x;\theta) = \frac{e^{\theta_i^T x}}{\sum_j^k e^{\theta_j^T x}}$

3. Decision Forest

It is one of the models quite commonly used in a lot of reviewed literature corresponding to stock trading predictions. The intent has been to analyze its perfomance compared to the other two models used.



For the ground truth labels, the 15-day future returns, calculated as percentage change on the Welles Wider smoothened closing prices, are used. The historical economical cycles are analyzed and identified to classify labels as strong-buy or sell in an economic uptrend, and as strong-sell or buy in an economic slowdown.

A profit/Loss calculation algorithm has been created and is used to automate the evaluation of the target labels performance. It is also used to compute the performance of the dev and test data.

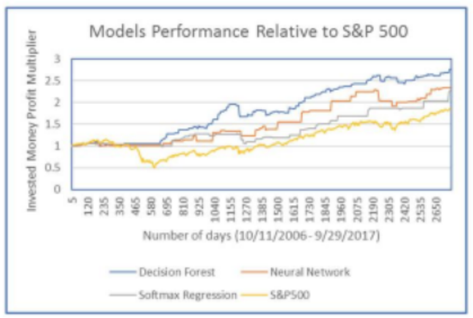
Results

For trading, the metric is profitability. The models' predictions for the buy and sell classes are used to generate medium/long term trading signals on the dev and test data. The resulting profit (or Loss) multipliers are tabulated in comparison to the S&P500 index performance over the same time period.

Two Trading Decision strategies are used. In the first strategy, the trading is triggered for all the four classes. In the second strategy, trading is triggered for the three classes - Strong Buy, Strong Sell and Sell.

Model & Trading Decision Strategy	Train Profit Multiplier	Dev Profit Multiplier	Test Profit Multiplier
Strong Buy, Strong Sell, Buy, Sell			
Neural Network	44.52	44.35	1.9
Softmax Logistic Regression	7.47	6.22	2.26
Decision Forest	105.21	65.03	2.15
Strong Buy, Strong Sell, Sell			
Neural Network	18.24	19.95	2.44
Softmax Logistic Regression	5.98	6.35	2.26
Decision Forest	40.5	24.79	2.77
S&P 500	3.95	3.82	1.86

The graph below shows the results on the test period from Oct. 2006 to Sept. 2017. Training and Dev used data from Jan. 1990 to Sept. 2006. The 80%-20% split was used between Training and Dev data.



The results look quite promising over the analyzed time period. This can be extended to the time frame encompassing all the available historical data, to further analyze the models' predictive quality.

Discussion

The economic indicators helped in identifying the beginning and end of recession periods. Different labeling techniques were tried, starting with two labels - buy and sell - on a daily basis. Such labels become hard for model to understand and correlate based on the feature set used. The data was then labelled based on prices and the over arching business cycle trends.
As the objective function has been to optimize stock returns, hence not all mistakes made by the model are equal. One misclassification can be more costly than another misclassification, Paint the target approach was applied in implementing a profit calculator to understand the model's performance in terms of returns target at the end of each run.

Future Work

The data set can be extended to start from 1900, to increase the train, dev, test data. The k-fold tecnique can be used to walk the test data overall a broad range of periods to validate the robustness of the models' prediction. Further optimization can be done by improving the cost function that penalizes the more severe misclassifications, and by using random forest for feature selection.This predictive framework can be extended to other stock market indexes, ETF's and individual stocks.

References

Introduction to Azure ML Studio
https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio
Azure ML Models
Neural Network: https://msdn.microsoft.com/library/azure/e8b401fb-230a-4b21-bd11-d1fda0d57c1f
Decision Forest: https://msdn.microsoft.com/library/azure/5e70108d-2e44-45d9-86e8-94f37c68fe86
Multiclass logistic: https://msdn.microsoft.com/en-us/library/azure/dn905853.aspx
Research papers
https://arxiv.org/pdf/1605.00003.pdf
http://www.sciencedirect.com/science/article/pii/S2405918815300179
Welles Wilders smoothing
https://www.tradingtechnologies.com/help/x-study/technical-indicator-definitions/welles-wilders-smoothing-average-wws/