Applying Machine Learning to Predict Stock Value

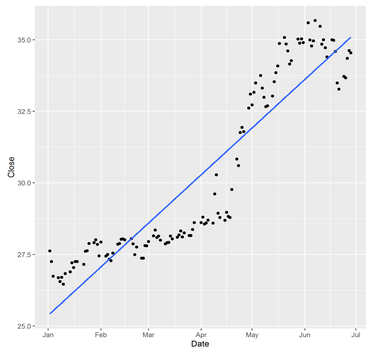
Joseph Lemley Yishui Liu Dipayan Banik Sadia Afroze

Faculty Mentor: Dr. Donald Davendra

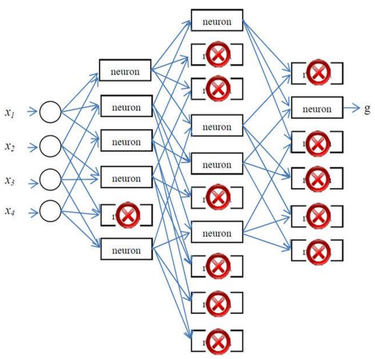
Abstract

The purpose of this study was to compare machine learning techniques for short term stock prediction and evaluate their effectiveness. Stock value analysis is an important element of modern economies. The ability to predict future stock prices from historical price values is of tremendous interest to investors. The prediction of stock performance is still an unsolved problem with a variety of techniques being proposed. Real stock values are affected by many elements, some of which cannot be measured. In this study, we limit our analysis to stock closing prices. We use these prices to predict the future stock value using regression and machine learning methods and compare their accuracy and effectiveness over a 3 year period.

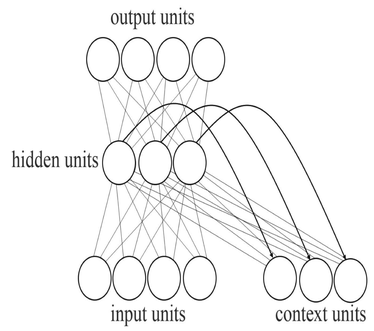
Methods



Linear Regression:
A linear regression model involves a line that is made to fit the data. The model is represented by two regression coefficients: the gradient and the intercept.
As Field[1] mentioned, the line with lowest sum of squared difference in all possible lines is the best fit.



GMDH (Group Method of Data Handing):
GMDH is a type of inductive algorithm with a self-organizing procedure. It finds the best fitting model by creating all combinations of inputs and choosing partial models with minimum external criteria (error calculated using different set of inputs) [3].
In this way, the GMDH algorithm can eliminate the need for guessing the size of the net work, and avoid over-fitting [2].



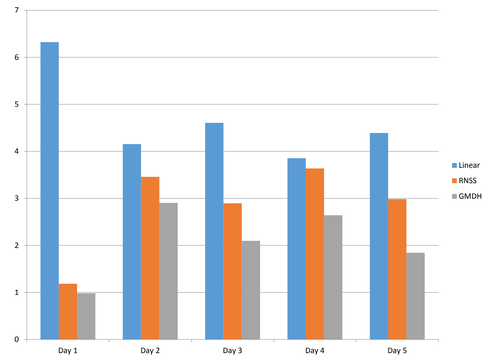
Elman Neural Network:
Elman neural Networks have an extra layer, called a context unit. This layer keeps the states of hidden units, such that when the next input is passed to hidden units, they still have access to prior states. This structure gives Elman networks an advantage in time series analysis [4].

Prediction

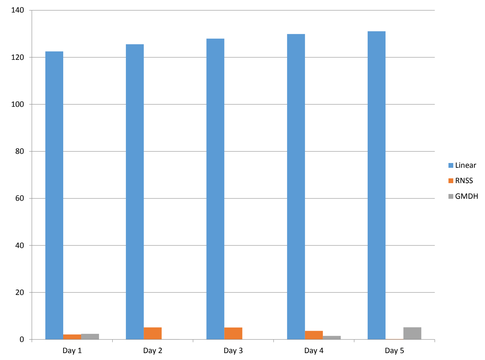
The three methods are used to perform a prediction using Microsoft and Apple stock closing prices in R using GMDH, the RSNNS library, and the lm() function. Functions take the data of a one year period between 2013 and 2015 from Yahoo Finance, and perform a prediction for the closing value of 5 market days after the last input date.

Then, the real closing value is collected and percentage error is calculated for each prediction, i.e. abs(prediction – real value) / real value. Part of the result is shown below.

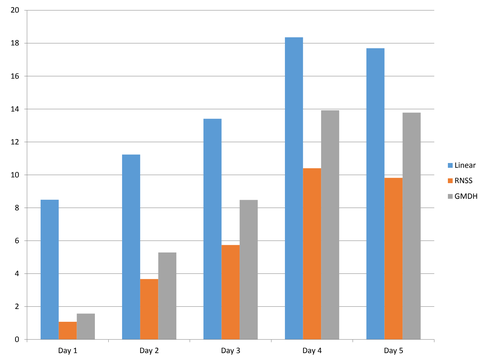
Percent of mean inaccuracy for 2014 for 5 day prediction period for MSFT.



Percent of mean inaccuracy for 2015 for 5 day prediction period for MSFT.



Percent of mean inaccuracy for 2016 for 5 day prediction period for MSFT.

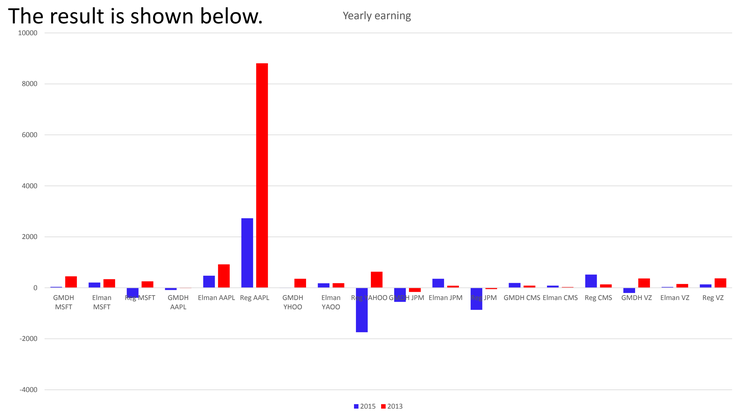


Benefit

From the previous section, it is obvious that for neural networks, next day prediction is more accurate. Thus, we designed a program that uses the last 150 days of data and rebuilds the model every market day.

To simulate an investment strategy that takes advantage of our predictions, we simulate the purchase of 10 shares at the beginning of every month, and if the prediction shows that the stock price will go up the next day, we buy 10 more shares. Otherwise, we sell all the shares we are holding. We continue this investment strategy for the entire month. At the end of every month, we sell all shares to see how much we earned for that month, and sum up the monthly income s to get the annual income achieved by following each model.

The result is shown below.



Conclusion

- Modern Machine learning approaches strongly outperform regression models.
- Elman RNNS slightly outperform GMDH when only closing prices are used for prediction.
- Closing prices are not enough to accurately predict stock prices, but they provide a significant source of information.

References

[1] A. Field, J. Miles and Z. Field, *Discovering Statistics Using R.* Thousand Oaks, CA: SAGE, 2012, pp. 318-320.

[2] “GMDH. “ Internet: http://www.gmhttp://www.gmdh.net/GMDH_abo.htmdh.net/GMDH_abo.htm [May 14, 2016].

[3] G. Onwubolu, *GMDH-Methodology and Implementation in C.* London: Imperial College Press, 2015, pp. 1-12

[4] J. L. McClelland. (2015, Dec. 16). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises.* [Online]. pp. 155-161. Available: https://web.stanford.edu/group/pdplab/pdphandbook/handbook.pdf