Stock Prediction with LSTM/SVR

Time-Series Financial Analysis with TensorFlow

Derrek Gass  
Engineering & Computer Science  
CSU-Sacramento  
 Sacramento, California USA  
 derrekgass@ecs.csus.edu

Alexander Lee  
Engineering & Computer Science  
CSU-Sacramento  
 Sacramento, California USA  
 alexwlee@ecs.csus.edu

ABSTRACT

There has been much development and interest in the application of machine learning to solve problems of pattern recognition in the analysis of signals such speech and images, and in recent years, has extended past this domain into many other fields. One such application is the understanding of patterns in financial time-series data. Based upon the past closing values for a stock, we would like to predict the value of the stock in the future with some degree of certainty so that traders may be informed about the state of their investment and respond with the proper course of action. Current methods rely on traditional financial indicators such as ARIMA, linear regression, moving averages, and other methods. For the purposes of risk assessment, anomaly detection, and optimization of asset portfolios, the implementation of machine learning proposes solutions to aide in this complex quantitative decision-making process. In this research project, we propose two different predictive models based upon previous research and developed with machine-learning tools; a support vector regression implementation, and a long short-term memory model. Both models performed with a RMSE of <0.05, and provide a baseline prediction of the next day value of an asset, indicating that further research may prove beneficial for solving the problem at hand.

CCS CONCEPTS

• Computing Methodologies • Applied Computing

KEYWORDS

Machine Learning, Artificial Intelligence, Finance, Computers in other domains

1 Introduction

• Briefly describe the problem you are addressing. What is the problem? Why is it important?

• Provide an overview of your approach. What is your approach?

• List your contributions. A bulleted list is fine.

• Show the organization of the rest of the paper.

2 Problem Formation

Precisely define the problem you are addressing (i.e. formally specify the inputs and outputs). Identify a well-known task or a dataset (e.g., a problem from Kaggle or UCI repository). Implement and evaluate at least two different models in terms of their performance on the selected task and dataset. Your goal here is to evaluate how well each model performs for your selected task. You must do feature normalization/encoding and parameter tuning.

3 System/Algorithm Design

Precisely define the problem you are addressing (i.e. formally specify the inputs and outputs).

3.1 System Architecture

Provide reader a high-level idea about your whole design/system. How all modules in your design interact with each other? This should talk about the tensorflow etc.

3.2 Support Vector Regression



Figure 1: SVR Network Architecture

*3.2.1 Support Vector Regression*

This algorithm is great in theory, yet it only is accurate for predictions about the next day. It can’t look farther ahead than one day with any kind of accuracy, so the model falls apart farther out from uncertainty. Minimizes error.



Figure 1: SVR Graphical Representation



Figure 2: SVR Formula

3.3 Long Short-Term Memory

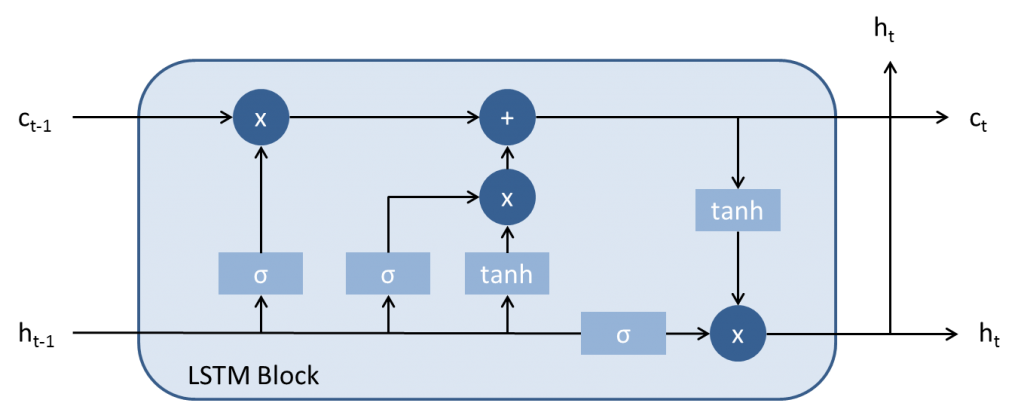


Figure 3: LSTM Overview

*3.3.1 Algorithm Description (Algorithm for Module 1)* This algorithm is great, except that it takes moment for the effect of the current day to have an impact in the model.



Figure 4: LSTM Layer

4 Experimental Evaluation

Precisely define the problem you are addressing (i.e. formally specify the inputs and outputs).

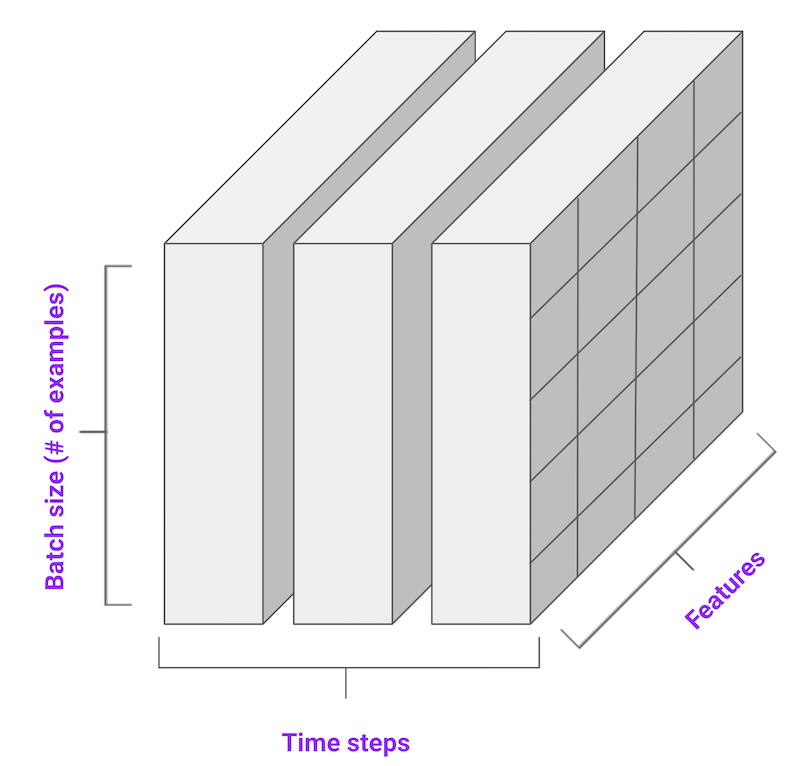


Figure 5: Encoding of Time-Series Data

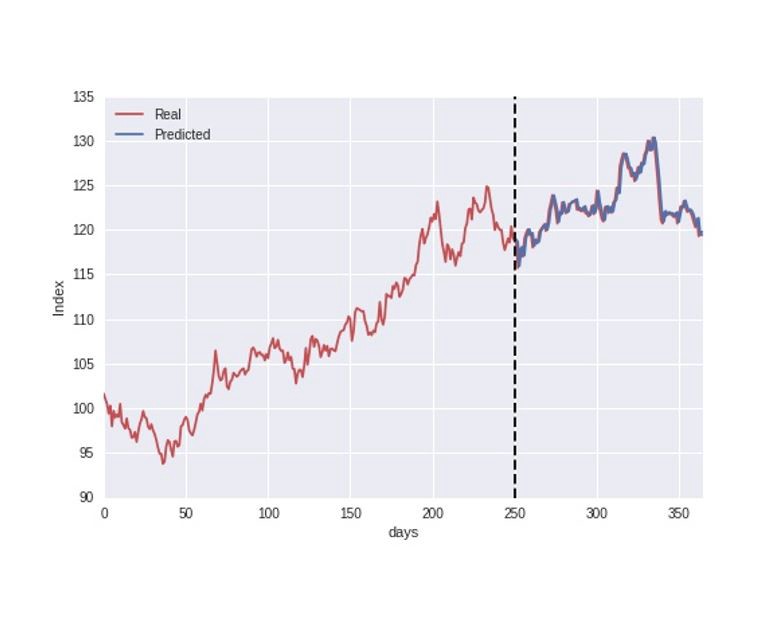
4.1 Methodology

• What was data used? How you split data to training and testing?

• What was the experimental setting?

• What metrics were used to compare different methods?

• What methods were implemented and compared? Make sure you include the competing methods that address the same problem as comparison baseline.

 Figure 4: Train/Test Split

1222 datum. Split 70 for train, 30% test.

the beginning of 2014 to the end of 2018.

X is a 36 x 1 array made up of the adj. closing prices of the last 36 days, y is the next day’s value.

Trained on 855 samples, validated on 367 samples.

20 epochs, batch size 10

Loss: mean squared error

Optimizer: adam

Dropout layers set to 0.2

One dense output layer predicting the next day’s asset value.

4.2 Results

Present the quantitative results of your experiments. Figures such as charts or histograms are frequently better than tables. For each figure, explain the result. What conclude we can draw from each figure?

5 Related Work

Answer the following questions for each related work that addresses the same or a similar problem.

• What is their problem and method?

• How is your problem and method different?

• Why is your problem and method better?

6 Conclusion

Briefly summarize the results and conclusions.

7 Work Division

A paragraph stating how the work is divided over all team members in your project.

8 Learning Experience

One or two paragraphs stating what you (and your partners) have learn from this project.

ACKNOWLEDGMENTS

Special thanks to Dr. Haiquan Chen. Without your guidance through the machine learning class, this endeavor would remain elusively out of reach to undergraduate students.

REFERENCES

[1] FNM Surname (2018). Article Title. Journal Title, 10(3), 1–10.

[2] F.N.M Surname, Article Title, https://www.acm.org/publications/proceedings-template.

[3] F.N.M Surname and F.N.M Surname, 2018 Article Title, The title of book two (2nd. ed.). Publisher Name, City, State, Country.

[4] Ian Editor (Ed.). 2018. *The title of book two* (2nd. ed.). University of XXX Press, City, Chapter 100. DOI: <http://dx.doi.org/10.1000/0-000-00000-0>.

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00