

COMPARISON of DIFFERENT KERNELS of SUPPORT VECTOR MACHINE for PREDICTING STOCK PRICES

Tanisha Shah ^{#1}, Iqra Shaikh ^{#2}, Ashishkumar Patel ^{#3}

[#] Computer Engineering Department, KSV University
LDRP-ITR, Sector - 15, Near KH - 5, Gandhinagar, Gujarat, India

¹ shah.tanisha22@gmail.com

² iqrashaikh617@gmail.com

³ ashish_ce@ldrp.ac.in

Abstract—The stock market prediction has always been a subject of interest since the emergence of prediction techniques. Because of the abundance of data available on the Internet in past years, it has been getting easier for all types of prediction. Various prediction and analysis models have been developed, majorly using the concepts of Machine Learning and Data Mining. Each model has its own advantage as well as limitations. Also, stock prices tend to fluctuate because of many factors like a national policy, or industry related political news. Many advanced analysis and prediction models do take these factors into consideration while some elementary models do not. In this paper, the analysis and its subsequent evaluation of stock prices are done using real-time data; based on the concepts of Support Vector Regression. Further, in this paper, the predicted stock prices are compared with the actual stock prices to have an idea regarding the accuracy of this technique.

Keywords - Stock Market Prediction, Support Vector Machine, Kernel, Radial basis function, Polynomial

I. INTRODUCTION

A. Stock Price

The stock price is highest or lowest amount of any company's stock which someone can buy. Stock prices of any company change every day. There are four values of any stock price. They are opening price of any stock on any particular day, closing price, highest price and lowest price. We have done analysis of past data for Apple company's opening price for July 2017. While prediction is also done for the opening price.

B. How prediction can be done

The efficient market hypothesis (EMH) states that stock prices can't be predicted [14]. But nowadays there are many approaches for forecasting stock price. In this paper, machine learning method is used. Machine learning is a very vast area. Machine learning is divided into 2 types- Supervised learning and Unsupervised learning [13]. In supervised learning, we are basically guiding the machine how to approach the task. While unsupervised machine learning is clustering where a bunch of random data is thrown to the machine and asked to form groups which fit best. We used supervised learning for our analysis for stock market prediction.

One such method of supervised machine learning is Support Vector Machine (SVM) [13]. SVM is mainly used for classification and regression. We have used the approach of regression in this paper. In simple language, regression means traveling back in time. Technically regression is an analytic method used to find the relationship between one or more independent variables and a dependent variable. Therefore, here the prediction of the stock market is done using this concept.

C. Support Vector Machine (SVM)

There are many methods which are effective as SVM like Artificial Neural Network (ANN). But there are few advantages of SVM over others [15]:

- 1) When high-dimensional spaces are there, then SVM is effective.
- 2) SVM has a simple geometric interpretation.
- 3) SVM always finds a global minimum meaning does not miss the bigger picture.
- 4) SVM are non-parametric thus their size is not fixed.
- 5) SVM is memory efficient.

SVM consists of SVR (it is the method) which has many kernels included in it. Thus various kernel functions make SVM versatile

II. LITERATURE SURVEY

In the past few years, many theories, as well as models, have been proposed and developed for Stock Market prediction.

Social media sites like Facebook, Twitter and even articles in newspapers and blogs have an effect on the fluctuation of stock market prices [8]. Many papers have extensively written regarding the impact of information available on Web using different data and text mining techniques. R.P. Schumaker and H.Chen [8] predicted stock prices and analysed various representation of the text written in newspapers and later compared them to Linear Regression. J.Bollen et al. [9] collected states of mood from Twitter feeds and analysed its text content using mood tracking tools. They found an accuracy of 87.6% in predictions of the stock market from the data collected.

Artificial Neural Network (ANN) model is another popular technique to predict the closing prices of stocks. J. T. Yao and C. L. Tan [10] wrote about a neural network prediction model built on seven steps which classify and predicts data. T.Hui-Kuang and K. Huarng [11] implemented a time series model to forecast stock prices of Taiwan. Neural networks can handle nonlinear relationships between numerical observations.

Md. R. Hassan and B. Nath [16] used Hidden Markov Model(HMM) to forecast stock prices. HMM is best suited to model dynamic systems and hence it is majorly used for pattern recognition and classification queries. The model that is chosen is trained on the past datasets yet it the prediction is not entirely accurate and straightforward.

III.METHODOLGY

When SVM is applied for just regression problem then it is termed as Support Vector Regression (SVR). The fact that SVR just minimizes the training error is a myth. It actually takes a stab at minimizing the generalization error.

SVR is a method of python for predicting stock prices. SVR is a method of library SVM. SVR method has many parameters. One such parameter is the kernel. The kernel in this method means the type of algorithm which is used to forecast the prices. SVR method in python has 5 types of kernels which mean 5 types of algorithms [16]. The various types are: Radial Basis Function (RBF), Polynomial, Linear, Sigmoid, precomputed or callable. Further, we have performed a comparison among RBF, Polynomial and Linear method.

A. Radial Basis Function

Radial basis function kernel is most common kernel function. It is the default function for SVR method in python. It is a non-linear regression. This means that data cannot be classified linearly. RBF is also known as the Gaussian kernel. RBF kernels are general purpose it should only be used when not using for text.

B. Polynomial

Polynomial is a kernel function used for support vector machines. The polynomial kernel is a non-linear regression. This regression is used only when data points are non-linearly separable. As the name suggests this kernel will have a polynomial equation. In python when we use polynomial kernel then another parameter needs to be set i.e. degree. This parameter specifies that how many degrees of the equation will be. By default, the 3-degree polynomial equation would be used if not mention. The space of polynomial kernel is same as polynomial regression. Polynomial regression is a type of regression analysis which is quite same as a linear regression but is modeled as nth degree instead of just 1 degree. It is also considered to be a variety of multiple linear regressions.

C. Linear

The Linear kernel is most commonly used for text. The Linear kernel is basically an equation which is defined as:

$$y = ax + b$$

where,

y = dependent variable

a = slope intercept

x = independent variable

b = constant

Here 'x' is variable which represents date and 'y' means stock price. Using x and y for past data 'a' and 'b' are obtained. After generating the equation, prediction can be made for any date by changing the value of 'x'.

Linear regression is concerned with only one independent variable. As the value of 'x' changes, we will get our value for 'y'.

D. Comparison

Fig. 1 represents comparison for 3 different types of kernels of Support Vector Regression. The graph is for the opening price of Apple Company. The black dots represent the opening stock price of Apple Company for different dates of July 2017. The red line represents the RBF kernel line used for prediction while green line represents linear kernel of SVR and lastly, the blue line represents polynomial kernel line. The red, green, blue dots represent the prediction made for 28th July using RBF, linear, polynomial kernel respectively for opening stock price.

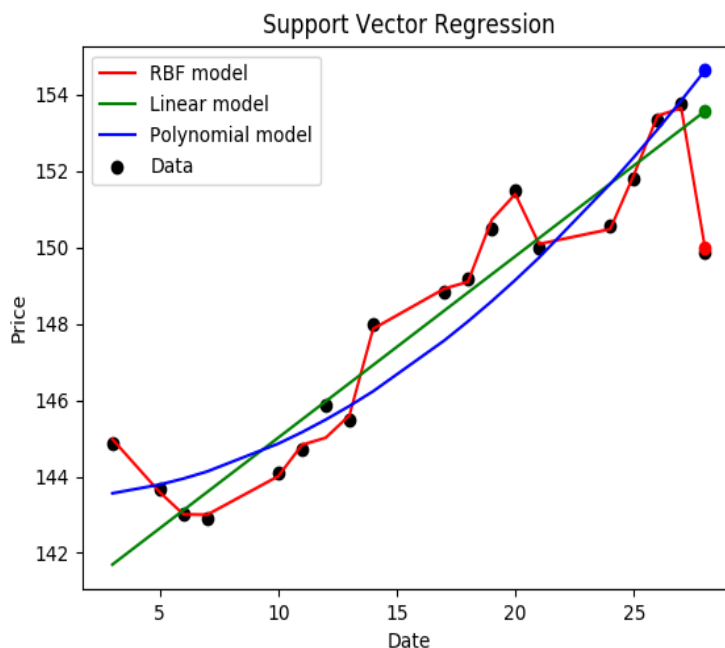


Fig. 1. Graph comparing 3 different methods on a single set of data.

It can be concluded from the graph(Fig. 1) that the prediction made for 28th July 2017 is more accurate to actual data for RBF kernel. The stock price for 28th July 2017 using RBF kernel is 149 which is very near to actual price which is 149.89(opening price). While using linear and polynomial the opening stock price is approximately 153 and 154 respectively.

To conclude, as RBF kernel is non-linear this means that it is non-parametric. Hence the complexity of the model is infinite. This means that more data used better and complex relationships could be determined. RBF is generally more flexible. Now as seen from the graph that data is not linearly separable thus higher accuracy could not be achieved. Despite being the simplest classifiers, linear kernels are better and faster in training and testing datasets as compared to RBF kernel. According to research linear kernel is mostly used when we want to optimize stock market prediction problem.

IV. CONCLUSION

Stock prices prediction has always been popular among researchers as well as investors because of the amount of financial risk involved. We conclude that for beginner analysts and brokers, support vector regression is an effective method to predict stock prices. But the stock market is dynamic in nature. Therefore, if an advance level accurate prediction is to make, then external parameters like natural disasters, political atmosphere, socio-economic conditions etc. have to be considered. Research and development have been done for implementation of several different techniques like Artificial Neural Network, Autoregressive Models, and prediction of stock prices using social media and news stories. However, in every case, there remains a scope for uncertainty and ambiguity as no technique is found to give an exact movement of stock prices. Hence these techniques can be used to reduce the losses incurred in stock market investments, but none of them guarantee a huge amount of profit for the same.

ACKNOWLEDGMENT

We would like to acknowledge our Head of Department Dr. Hiren Patel and the Computer Engineering department of LDRP-ITR, Gandhinagar for their continuous encouragement and support and for providing the necessary guidance that made this paper possible.

REFERENCES

- [1] P. K. Sahoo, K. Charlapally, "Stock Price Prediction Using Regression Analysis", International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015.
- [2] S.P. Das, S. Padhy, "Support Vector Machines for Prediction of Futures Prices in Indian Stock Market", International Journal of Computer Applications, Volume 41– No.3, March 2012.
- [3] M. Halls-Moore, "Support vector machines: A guide for beginners," September 2014. [Online]. Available: <http://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners>
- [4] Han Lock Siew, Md Jan Nordin, "Regression Techniques For The Prediction Of Stock Price Trend", 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE).
- [5] J. Zhang, "Applying time series analysis builds stock price forecast model," Modern Applied Science, vol. 3, no. 5, 2009.
- [6] R.V. Argiddi, Dr. Mrs. S.S. Apte, B. U. Kale, "An Analysis on Stock Market Intelligence and Research Approaches", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 1, January 2014
- [7] S.A.S. Olaniyi, Adewole, Kayode S., Jimoh, R.G. "Stock Trend Prediction Using Regression Analysis – A Data Mining Approach", ARPN Journal of Systems and Software, Volume 1 No. 4, JULY 2011
- [8] Robert P. Schumaker and Hsinchun Chen. "Textual analysis of stock market prediction using breaking financial news: The AZF in text system" ACM Trans. Inf., 2009.
- [9] Johan B, H. Mao and X. Zeng. "Twitter mood predicts the stock market" Journal of Computational Science, 2011.
- [10] JingTao YAO and Chew Lim TAN, "Guidelines for Financial Prediction with Artificial neural networks",
- [11] Tiffany Hui-Kuangyu and Kun-Huang Huang, "A Neural network-based fuzzy time series model to improve forecasting", Elsevier, pp: 3366-3372, 2010.
- [12] Md. Rafiul Hassan and Baikunth Nath, "Stock Market forecasting using Hidden Markov Model: A New Approach", Proceeding of the 2005 5th international conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.
- [13] Galit Shmueli, Nitin R. Patel, Peter C. Bruce, "Data mining for business intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner", 2nd ed.
- [14] Qaiser Munir, Sook Ching Kok, "Information Efficiency and Anomalies in Asian Equity Markets: Theories and evidence"
- [15] 2011, "Scikit-learn: Machine Learning in Python" website [Online] Available: <http://scikit-learn.org/stable/modules/svm.html#support-vector-machines>.
- [16] 2011, "Scikit-learn: Machine Learning in Python" website [Online] Available: <http://scikit-learn.org/stable/modules/generated/ssklearn.svm.SVR.html>

AUTHOR PROFILE

Tanisha Shah is currently pursuing computer engineering (4th year) at LDRP Institute of Technology and Research a constituent institute under Kadi Sarva Vishwavidyalaya

Iqra Shaikh is currently pursuing computer engineering (4th year) at LDRP Institute of Technology and Research a constituent institute under Kadi Sarva Vishwavidyalaya

Mr. Ashish Kumar B. Patel working as an Assistant Professor at LDRP Institute of Technology and Research a constituent institute under Kadi Sarva Vishwavidyalaya from last 11 years. His area of research is Data Mining in which he is pursuing Ph.D. in Recommender System.