Stock Market Prediction

Amarapalli-Bavirisetty-Gudivada-Jain

University of Cincinnati

CS 6037: Machine Learning

1. Abstract:

Stock Market is a complex system and is very difficult to analyze the impacting factors before coming to a decision. One decision can make huge impact on an investor's life. So, predicting the future stock value will benefit the investors greatly.

In this project, we designed a Machine Learning model which will predict the future value of a stock from the current data. Here, we used the stock data (750 instances) of each week for 6 months, January to June 2011. This data was used to test and train our algorithm. We predicted the opening stock price over the following week. We implemented this model using different algorithms and techniques. We analyzed the results and identified the best performed algorithm.

2. Introduction:

The Dow Jones Industrial Average (DJIA) is a stock market index created by Wall Street Journal editor and Charles Dow (Dow Jones & Company co-founder). DJIA is a collection of thirty publicly traded stocks. These include multi-national companies like Apple, Coca-Cola, Intel, etc. The DJIA value is obtained by the compilation of values of each stock. The DJIA value helps to find the strength of the United States economy among other things. The closing value is reported on news networks across the United States. The sum of the prices of all the 30 stocks is divided by the Dow Divisor. The Dow divisor on September 1, 2017 was 0.145233968773. [1]

3. Problem:

Stock market is a system where the prediction of the stock market will be helpful to the investors. The investors decision to purchase a stock will change their lives in a broad way. Consider the case, where the decisions made by the investors could never go wrong and can predict the price of the stock for the next day, week or even a month. If this is the case, then many investors would become millionaires by choosing what stocks to purchase. They purchase stocks for a low price and try to sell at a high price. This is the case which motivates our team to

consider the problem of predicting the stock market.

4. Project Goal:

Project Goal is to predict which stocks in the DJIA will have a net value gain over the next week given data for the current week. This goal helps the investors to choose what stocks to purchase and make confident and safe decisions in investment.

5. Description of Data set:

The Dataset chosen incudes 750 instances of Dow Jones stock and is retrieved from the UCI Machine Learning Repository [2]. The performance of every DJIA stock in the first and second financial quarters of 2011 is represented by the dataset. The features in the dataset include the stock symbol, the weekly opening, high, low, and close prices for the stock etc. Based on financial quarter the data is partitioned into two sets, 360 examples for first quarter and 390 examples for second quarter. The Stock Symbol in the data is in text format and cannot be used directly to train the model. Hence, each stock symbol is assigned an index from 1 to 30, to make the data accessible. The entire data, which is in string format is then converted into number format. The missing values in the data are filled with average of the feature for the appropriate Stock Symbol. The following 10 features were used to train and test the data:

1. Quarter
2. Stock Symbol
3. Opening price for the week
4. Highest value of stock for the week
5. Lowest value of stock for the week
6. Closing price for the week
7. Volume of shares traded
8. Percentage change in price over the previous week
9. Percentage change in volume over the previous week
10. Previous week's volume

6. Approach:

Initially, we partitioned the dataset into two random halves. First half data is used as training set, while the second half is used as a test set. The feature which is insignificant i.e. which has the least effect on performance of the model is eliminated using L2-Regularization. Using the 3-fold cross-validation on the whole data set, the performance for degree 1 and degree 2 are evaluated. As the results for degree 2 were better, further training was carried out with degree 2.

6.1 Algorithms:

The following Regression Algorithms were tested:

- *Linear Regression:* Linear Regression is perhaps one of the most familiar

algorithms in machine learning. It is a linear approach for modelling the relationship between the input and output. Here we approached in two different ways, namely, Normal Equation method and Gradient descent method, which vary in the process of finding the weights vectors.

- *Support Vector Regression:* Support vector machine (SVM) analysis is one of the popular machine learning tools for regression and classification analysis. It is considered as a nonparametric technique because it relies on kernel functions. Here the default function Gaussian kernel is used.
- *Regression trees:* Regression trees are the decision trees where the target variable can take continuous values. So, we used regression trees for this regression problem.

7. Results:

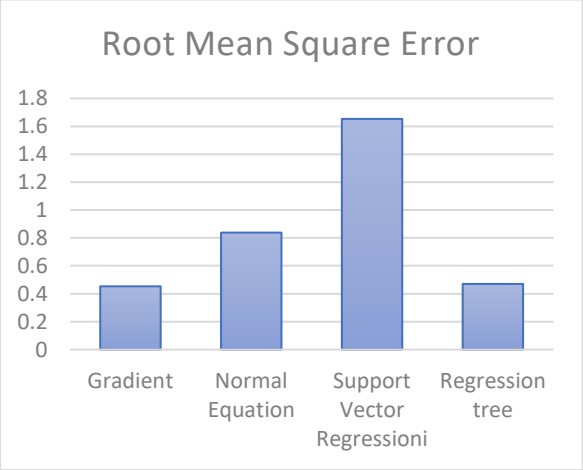

Figure 1: Root Mean Square Errors for each Algorithm

Once the insignificant feature is identified and eliminated, we used the data set of degree 2 for further training of the model. We used the Train dataset to develop the model and tested it on the Test dataset. The Root Mean Square error for each algorithm was computed and plotted in the below graph.

8. Conclusion:

Based on the above results and graphs plotted for Sample test data, it is evident that regression tree and gradient descent algorithms are best suitable for this regression problem.

We believe that, the error in prediction can be further decreased by training the model with a wide range of data, probably from a different year, a different quarter and then evaluate the output for each algorithm and determine if the chosen model can be implemented for the problem.

9. References:

[1] http://online.barrons.com/mdc/public/page/9_3022-djiahourly.html

[2] https://archive.ics.uci.edu/ml/datasets/dow+jones+index

[3] Machine Learning by Tom M.Mitchell

[4] https://www.mathworks.com/