Forecasting the Direction and Strength of Stock Market Movement

Jingwei Chen cjingwei@stanford.edu

Ming Chen mchen5@stanford.edu

Nan Ye nanye@stanford.edu

Abstract - Stock market is one of the most complicated systems in the world, and it has connection with almost every part in our life. In this paper, we applied different Machine Learning methods to predict the direction and strength of the market index price movement. Specifically, we looked at Multinomial Logistic Regression, K-nearest neighbors algorithm, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Multiclass Support Vector Machine, and compared their performance results based on test accuracy, robustness, and run-time efficiency.

1. Introduction

Stock market movement prediction is a challenging task because of the high data intensity, noise, hidden structures, and the high correlation with the whole world. In addition to forecasting the movement prediction, we also tried to predict the movement strength of stock market at the same time.

We started with a brief description of the response and the predictors, which we think are relevant, and did feature selection based on the correlation coefficients between them. Finally, we applied different Machine Learning methods to train the data and make predictions. The data is from Yahoo! Finance and Federal Reserve Bank of St. Louis.

2. Dataset

The response in our models is based on the historical data of S&P 500 Index. We divided the returns of S&P 500 Index into four categories based on our "segment threshold" δ as following: $(-\infty,-\delta],(-\delta,0],(0,\delta],(\delta,+\infty)$.

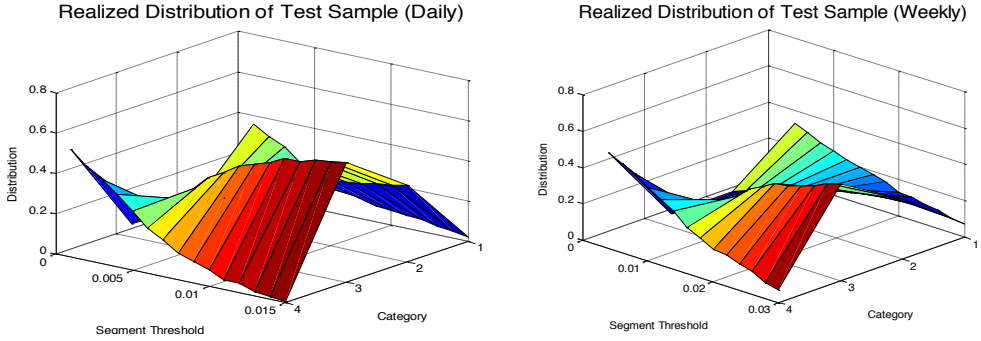
We marked them with Class 1, 2, 3 and 4, so that we can integrate the information about stock market movement and its strength together. Our predictors include the lagged change rates of following indices - S&P 500 Index, US Dollar Index, Volatility Index (VIX), FTSE 100 Index, Nikkei 225 Index, Hang Seng Index, Shanghai Stock Exchange Composite Index, S&P 500 Index Volume, S&P 500 Index highest price change rate during the period, S&P 500 Index lowest price change rate, 3-Month T-Bills Rate, 10-Year T-Notes Rate and Initial Jobless Claims. The range of the data is from 12/28/1990 to 11/1/2013. Therefore, there are 5758 observations in the daily data and 1158 observations in the weekly data.

3. Evaluation Metric

We set aside the last 200 observations in our data sets as the test data because we only care about the predictability of our models in the recent period and the best model might change as time flies. We trained our models on the data based on different training dataset size and different "segment thresholds" and then applied the coefficients we obtained to make predictions on the test data. We use test precision defined as $\sum_i 1\{y_{pred}^{(i)} = y_{actual}^{(i)}\} / \sum_i 1\{y_{actual}^{(i)}\}$, where $y_{pred}^{(i)}$ and $y_{actual}^{(i)}$ are the i-th predicted response and the i-th actual response. Also, we looked at the test precision based on the probability of Class 4 occurring when we predict the response is 4, i.e.,

$$P(Y_{actual} = 4 \mid Y_{pred} = 4) \approx \frac{\sum_i 1\{y_{pred}^{(i)} = y_{actual}^{(i)} = 4\}}{\sum_i 1\{y_{pred}^{(i)} = 4\}}.$$

The realized distributions of our test set on the 4 categories are shown below.



4. Multinomial Logistic Regression

Multinomial Logistic Model with four categories has three logit functions:

$$Logit\ Function\ for\ Y = i\ relative\ to\ logit\ function\ for\ Y = 4, \qquad i = 1,2,3$$

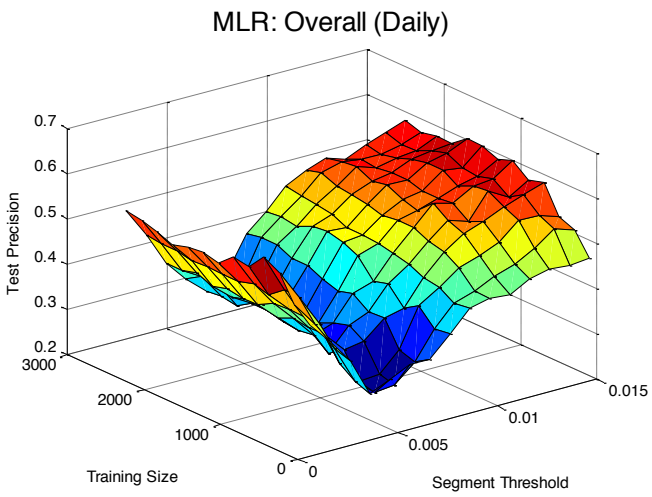
while Class Y = 4 is the reference group. So we can write the logit functions as the following:

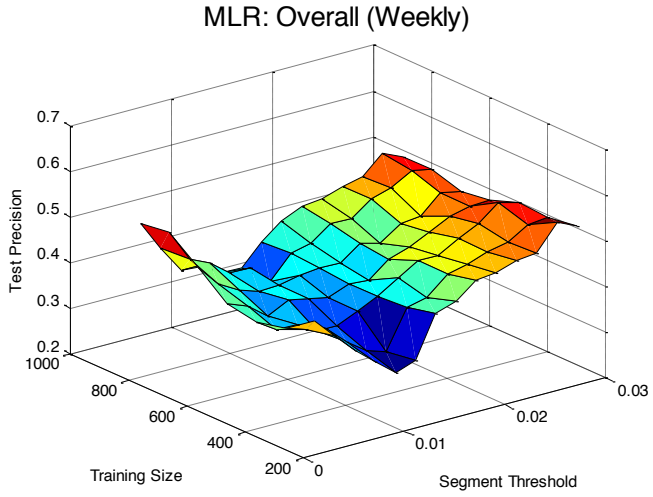
$$log\big(g(Y=i)\big) = A_1 + B_{i1}X_1 + \cdots + B_{ik}X_k, \qquad i = 1,2,3$$
$$log\big(g(Y=4)\big) = log1 = 0,$$

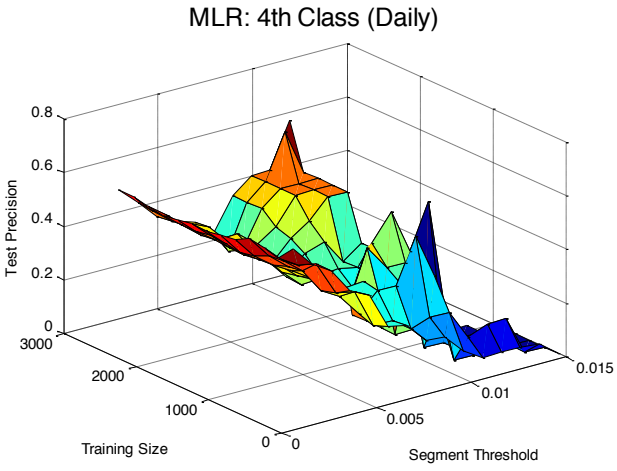
where A_i and B_{ij} are the regression coefficients. The Multinomial Logistic Functions are then defined as:

$$f(Y=i) = \frac{g(Y=i)}{g(Y=1)+g(Y=2)+g(Y=3)+1}, \qquad i = 1,2,3$$
$$f(Y=4) = \frac{1}{g(Y=1)+g(Y=2)+g(Y=3)+1}.$$

We then only need to compare the four Logistic Functions and pick the greatest one as the predicted category. The out-of-sample category prediction results based on daily and weekly data are shown below.





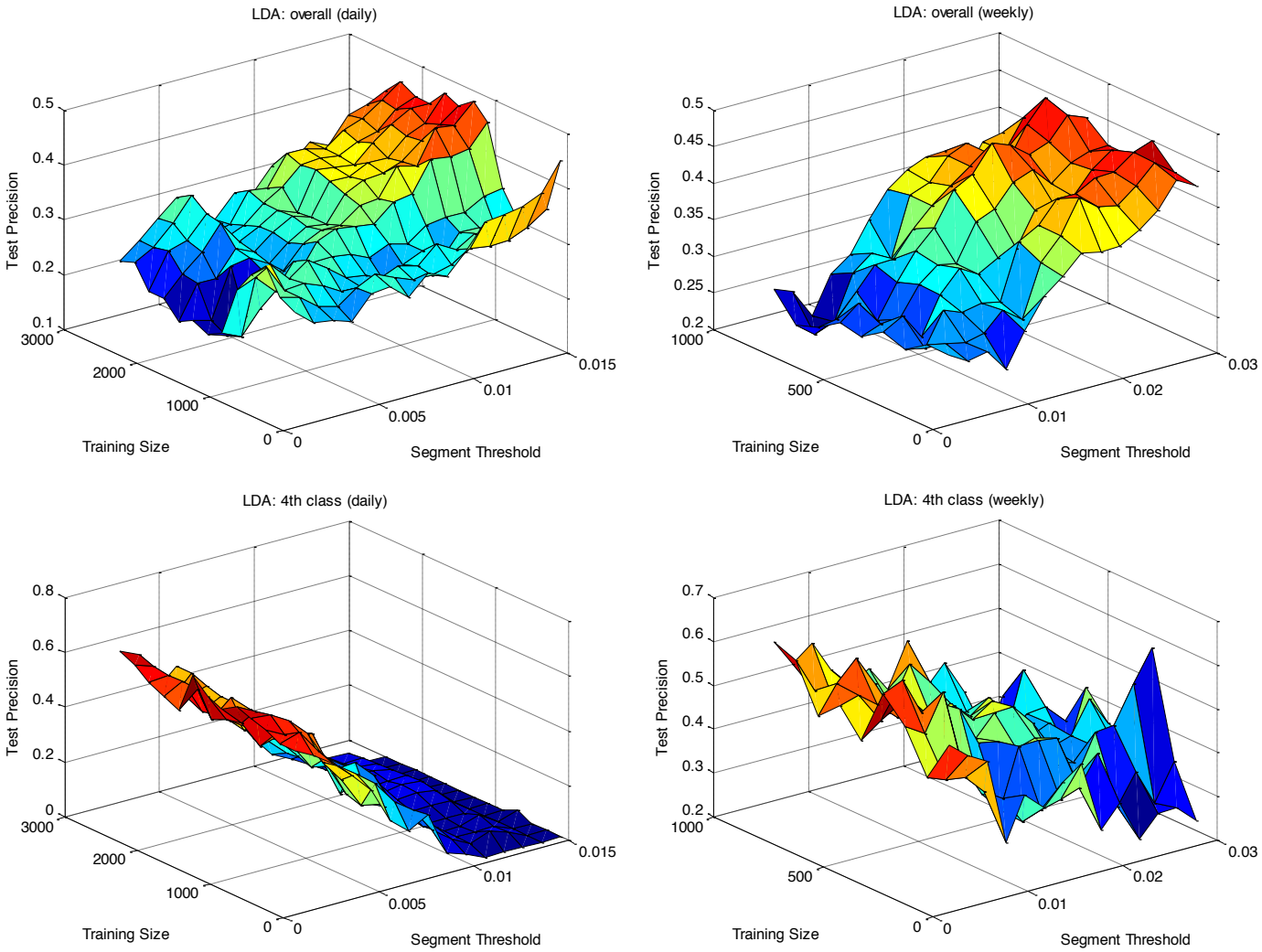


The first row of graphs presents the overall category prediction results based on our MLR model. As the segment threshold increases (above 0.5% for daily and 1% for weekly), both daily and weekly data perform better; however, the performance of weekly data is less volatile than the daily data, and weekly data generally yields a better performance with respect to the same level of realized distribution of the test sample we use.

The situation in the second row is similar to that of the first row. In terms of the 4th Class (i.e. Y = 4 or Extreme Positive), weekly-data-based model significantly outperforms the one based on daily data with respect to both accuracy rate and reliability. Therefore, weekly-data-based MLR model is more prosperous and reliable for real-world real-time prediction.

5. Linear Discriminant Analysis

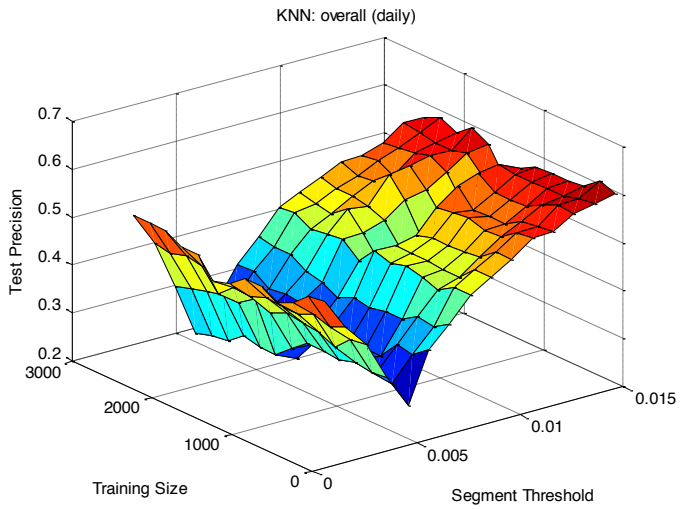
We applied both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) to model the data. QDA assumes that each class has its own covariance matrix. Consequently, LDA has lower variance, which leads to improved prediction performance. As we performed both models, we found out that LDA substantially outperforms QDA. Thus, we will only present the LDA model here. With different values of training size and segment threshold, the overall and 4th Class test precisions with weekly and daily data are shown as below.

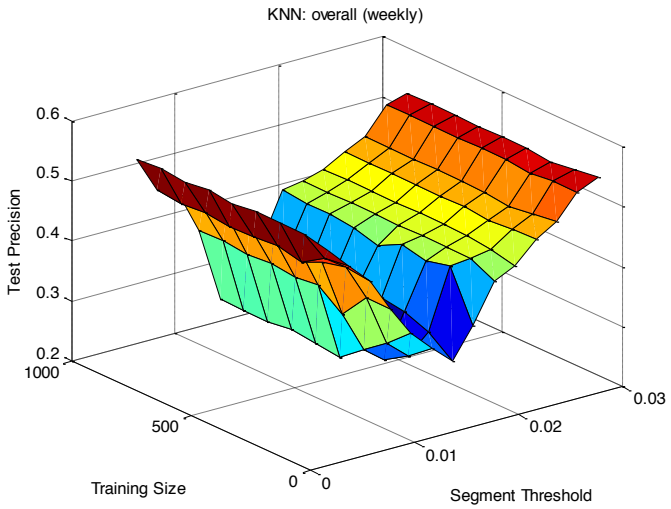


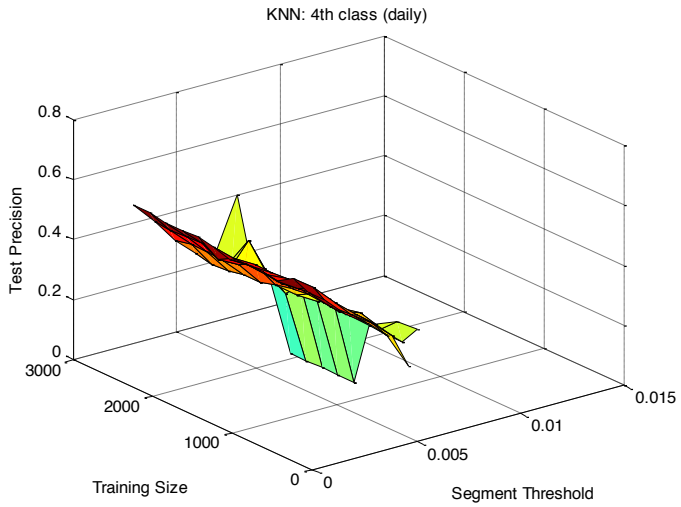
As can be seen from the figures above, weekly data performs better than daily data and the test precision does not change significantly with different values of training sizes. However, it severely depends on the segment threshold value. It is worthy to note that, when threshold becomes larger, the observation in Class 3 will increase and our model tends to predict Class 3, thus the overall precision increases while the 4th Class precision drops.

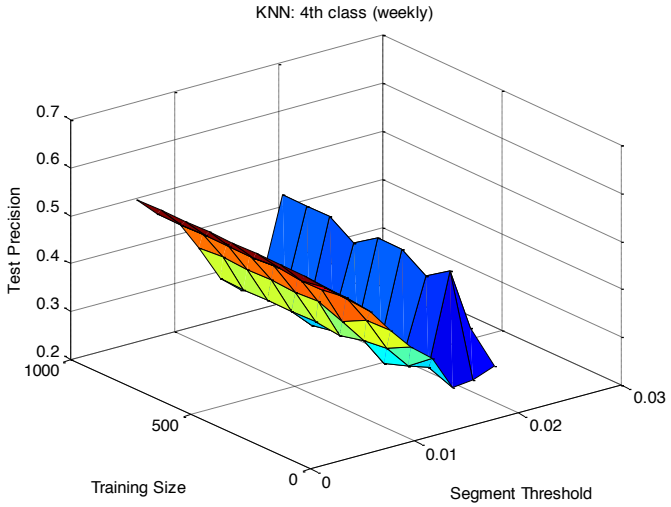
6. K-Nearest Neighbors Algorithm

We performed sensitivity analysis with different values of K and it turned out that K=100 would be a reasonable choice. With different values of training size and segment threshold, the overall and 4th Class test precisions with weekly and daily data are shown below.







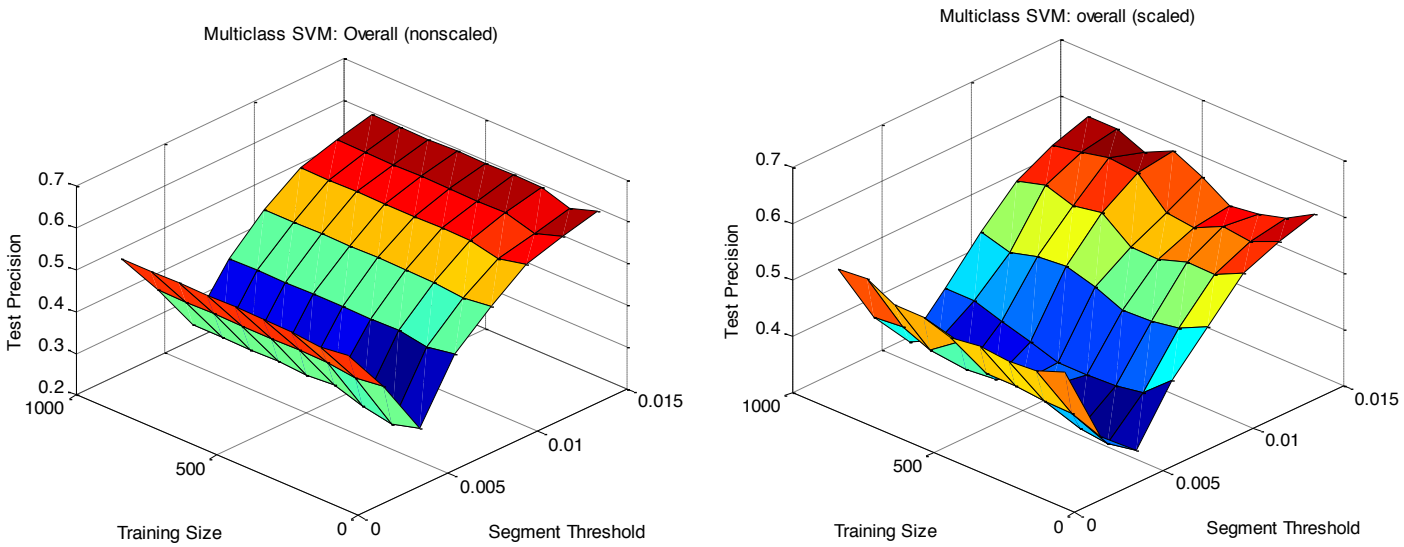


We can see that KNN model underperforms LDA model and shows bad results for medium segment threshold values, nor is its prediction accuracy for 4th class satisfying. Besides, the running speed is also much lower than LDA. Therefore, we recommend LDA to model data.

7. Multiclass Support Vector Machine

The last model we used is the Multiclass SVM Model, which consists of the preceding observations of the test point. Because we have 200 test points, so we should train our models for 200 times and make predictions separately. We applied the same "segment thresholds" to categorize the returns of S&P 500 Index, except that we changed the range and the size of step. We trained our models on the scaled and non-scaled daily data and weekly data, and we also applied different kernels to the model, including linear kernel, polynomial kernel, radial basis function kernel, and sigmoid kernel.

After trying out all the kernels above, however, we discovered that the linear kernel and the polynomial kernel do not apply to this problem, and sigmoid kernel behaves similar to radial basis function kernel. Moreover, the outcome of daily data is similar to that of weekly data. Therefore, we only show the results of our SVM models using radial basis function kernel on daily data. The results of our Multiclass SVM are shown below.



We can see that scaled predictors perform more robust than non-scaled predictors because the range of test precisions is narrower on scaled predictors. Also, we can conclude as before that the training size has little effect on the test precision while "segment threshold" affects the test precision significantly. In addition, we can achieve more than 60% test precision if we choose a proper "segment threshold" no matter whether we scale the predictors.

8. Conclusion

Based on the results from different machine learning methods, we can see that Multinomial Logistic Regression performs the best in all the models in terms of the model robustness, prediction precision, and run-time efficiency. All the models perform better on the weekly data than on the daily data. For Multiclass SVM, scaled predictors can improve the robustness of the model in the prediction.

Reference

G.James, D.Witten, T.Hastie, R.Tibshirani (2009). An Introduction to Statistical Learning. Springer.

R.Choudhry, K.Grag (2008). A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology, 15*.

W.Huang, Y.Nakamori, S.Wang (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research, 32*, pp. 2513–2522.