Predicting the Trend of Change of Foreign Exchange Rate Using Support Vector Machine

Syuan-Ming Guo

INTRODUCTION

A foreign exchange (FX) rate is the amount of one currency (the base currency) that must be sold (bought) in order to buy (sell) one unit of another currency (the counter currency) [2]. Correctly predicting the trend of foreign exchange rate is the key of profitable trading. In this project, I test the ability of support vector machine (SVM) to predict the direction of change the exchange rate based on solely the historical rate. The foreign exchange rate data used in this project are EUR/USD exchange rate from 2002/07/01 to 2010/11/04 with time interval = 15 minutes, and EUR/USD exchange rate from 1999/01/04 to 2010/11/04 with time interval = 1 day. The data consist of a sequence of four representative prices in a given time interval: opening, closing, highest, and lowest. The results of this project demonstrates that SVM can be used to predict the trend of change of future highest and lowest currency exchange prices with decent performance (~35% error rate).

METHODS

Features and labels

Foreign exchange rate data are time series data. For the k-th time interval, I used the four reprehensive prices before the k-th time interval (k-1 or earlier) to construct the features, and the price information from the k-th time interval to define the label. The features I used for classification include basic features such as the changes (or first derivatives, velocities) of the four representative prices, and the moving averages of velocities with different window size and weight (uniform or exponential). In addition, I included several indicators commonly used in foreign exchange trading as features. The indicators I included are: Moving Average Convergence Divergence (MACD), stochastic oscillator (STO), and Relative Strength Index (RSI). The label can be defined by the trend of one of the four representative prices from k-1 to k-th time intervals (+1 for rising, and -1 for dropping or no change). For this project, I tested three cases where labels are determined by the changes of highest, lowest, and closing, respectively.

SVM Classifier

I used the MATLAB code provided by MIT 6.867 (Machine Learning) to build my SVM classifier. Importantly, features were normalized before used to train the classifier. I scaled binary feature vectors to [-1,1] and continuous feature vectors to standard deviation of 1. The normalization procedure significantly improved the performance of the SVM classifier.

To find the optimal classifier with the highest accuracy, different kernels and combination of parameters are screened. For radial basis kernel (or Gaussian kernel), a grid search is performed for the standard deviation of the Gaussian function D together with the penalty for constraint violation C. For polynomial kernel, the degree of the polynomial D together with the penalty for constraint violation C are varied.

Cross-validation was implemented by using part of the whole dataset as training set and left out the rest as test set, and then repeated this procedure for 10 times with different choices of the training set. The performance of classifier was evaluated by the averaged error rate of the cross-validation. I included 200 time points for each training set. Larger size of training set was also tested on radial basis kernel but did not improve the performance significantly.

RESULTS AND DISCUSSIONS

Exchange rate in short time interval

First, I trained the SVM classifier using the EUR/USD exchange rate from 2002/07/01 to 2010/11/04 with time interval of 15 minutes. After feature and label construction, there are 199759 usable data. In the case where the label is defined as the change in the highest (or the lowest) price, the pair-wise feature plot matrix shows different labels might be separable for some features (e.g. changes of the closing and the lowest prices) (Figure 1). However, in the case where the label is defined as the change in the closing price, no features seem to be able to separate different labels well (Figure 2). This is mostly likely due to that the time spanned by the highest (or lowest) price from k-1 to k-th time interval is partially overlapped with the time spanned by the closing price from k-2 to (k-1)-th time interval. The partial overlap time spans cause higher correlation between the price changes and thus makes the labels more separable by the features when the label is defined as the change in the highest (or the lowest) price. In contrast, the time spanned by the closing price from k-1 to k-th time interval is not overlapped with the time spanned by any price from k-2 to (k-1)-th time interval (as closing price is the final price in each interval). The lower correlation

between the price changes makes the labels less separable by the features when the label is defined as the change of the closing price (Figure 2).
The classifier performance is also consistent with the pair-wise feature plot matrix. Grid search of parameter shows the minimal error rate for predicting the trend of the highest price is ~34% (Figure 3). The minimal error rate for predicting the trend of the lowest price is ~33% (Figure 4). Minimal error rate using the polynomial kernel and Gaussian kernel are similar. However, the minimal error rate for predicting the trend of the closing price is ~47%, which is only barely better than random guess.

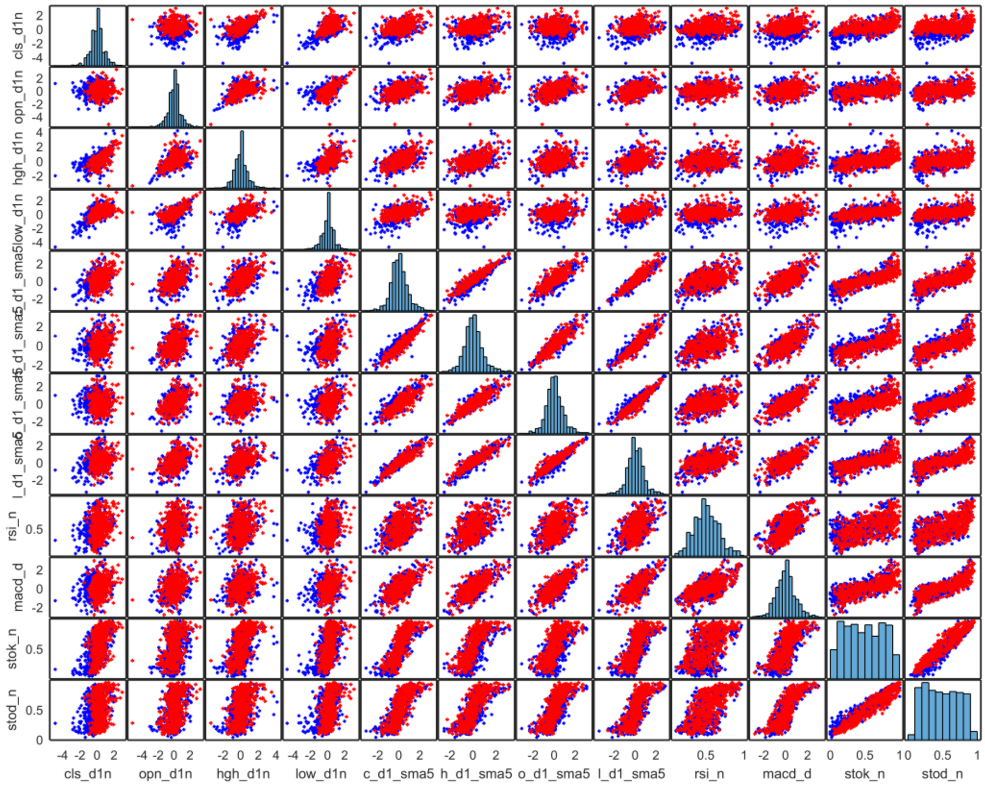

Figure 1: Pair-wise feature plot matrix for 15-minute exchange rate data. The points have label +1 (red) if the highest exchange rate from t-1 to t increases, and have label -1 (blue) if the highest exchange rate from t-1 to t decreases or remains the same.

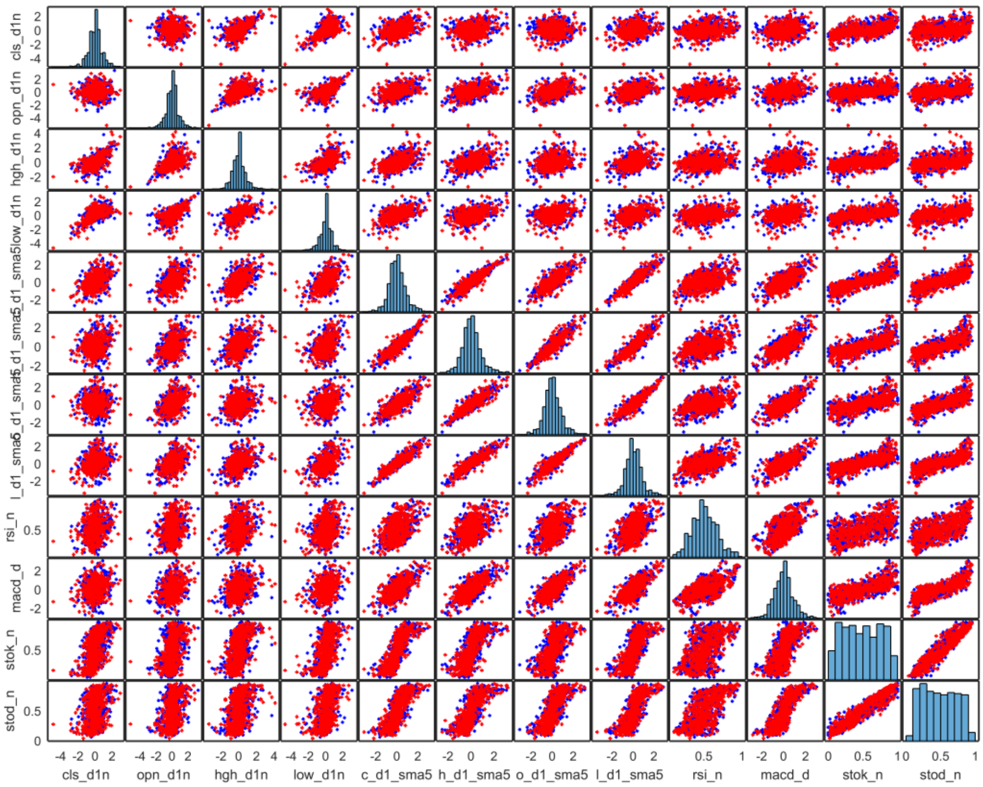

Figure 2: Pair-wise feature plot matrix for 15-minute exchange rate data. The points have label +1 (red) if the closing exchange rate from t-1 to t increases, and have label -1 (blue) if the closing exchange rate from t-1 to t decreases or remains the same.

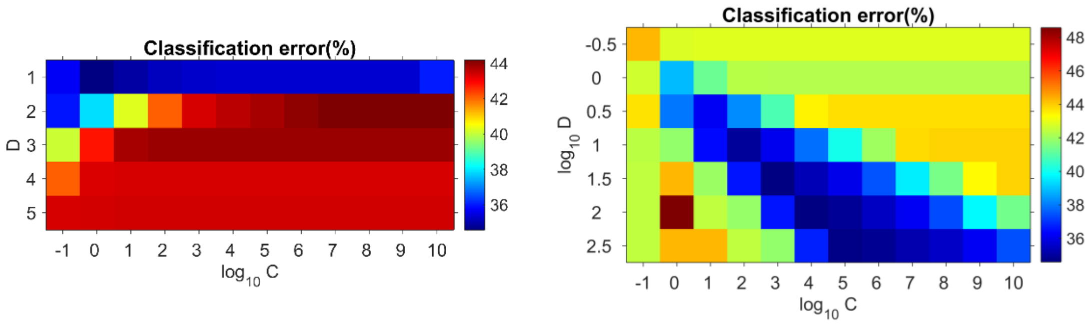

Figure 3: Average error rate of the classifier to predict the trend of the highest exchange rate from t-1 to t for 15-minute data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

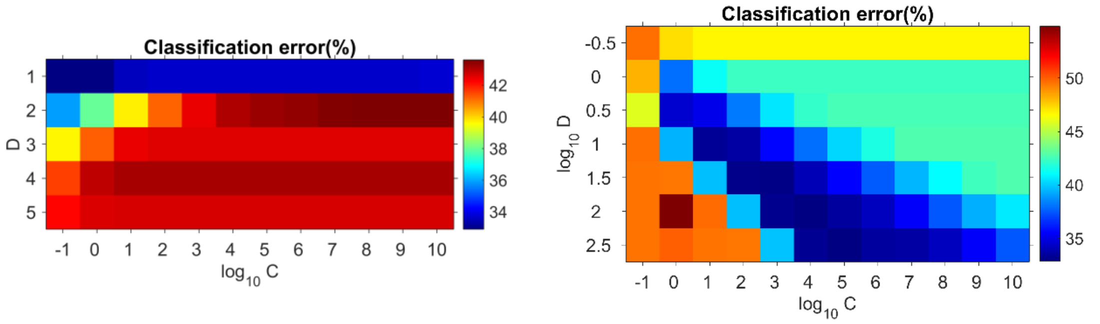

Figure 4: Average error rate of the classifier to predict the trend of the lowest exchange rate from t-1 to t for 15 minutes data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

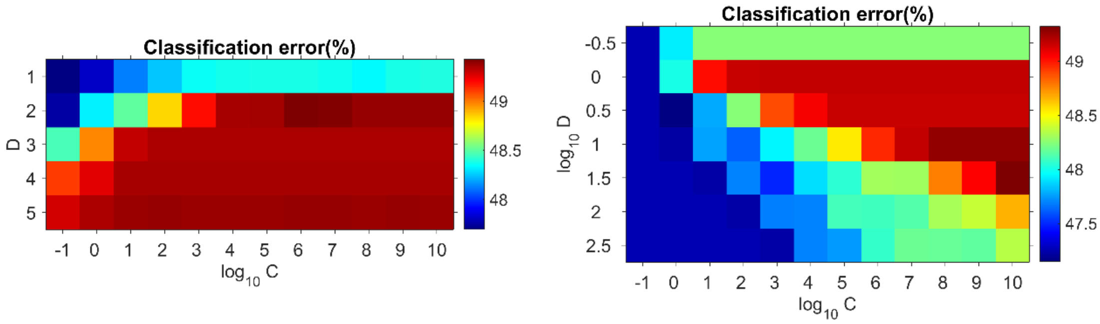

Figure 5: Average error rate of the classifier to predict the trend of the closing exchange rate from t-1 to t for 15 minutes data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

Exchange rate in long time intervals

Next, I trained the SVM classifier using the EUR/USD exchange rate from 1999/01/04 to 2010/11/04 with time interval of 1 day. After feature and label construction, there are 2770 usable data. Results of long time interval data are pretty similar to the results of short time interval, suggesting the correlation between features and labels are largely invariant with the time interval (Figure 6-Figure 10).

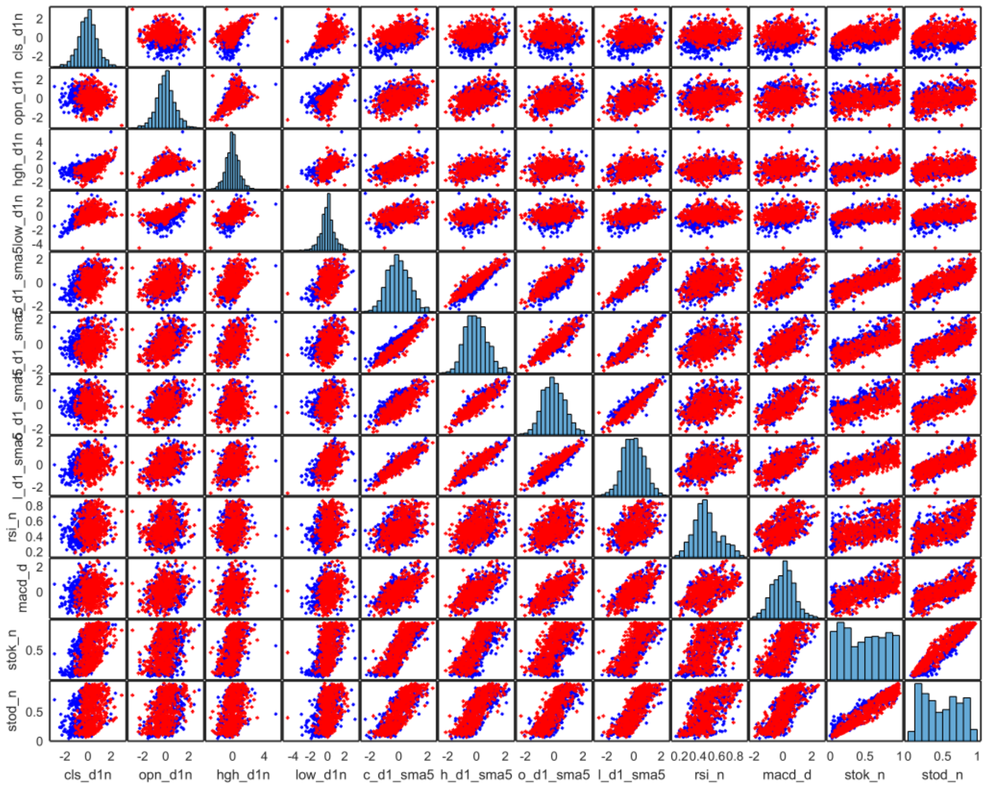

Figure 6: Pair-wise feature plot matrix for daily exchange rate data. The points have label +1 (red) if the highest exchange rate from t-1 to t increases, and have label -1 (blue) if the highest exchange rate from t-1 to t decreases or remains the same.

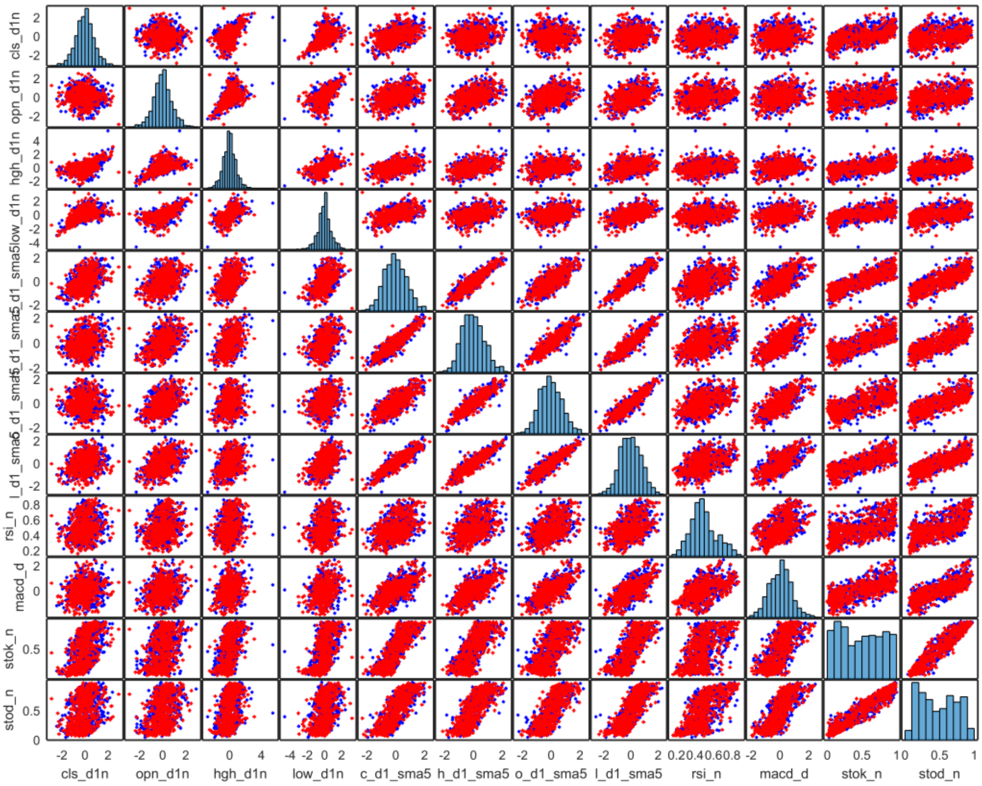

Figure 7: Pair-wise feature plot matrix for daily exchange rate data. The points have label +1 (red) if the closing exchange rate from t-1 to t increases, and have label -1 (blue) if the closing exchange rate from t-1 to t decreases or remains the same.

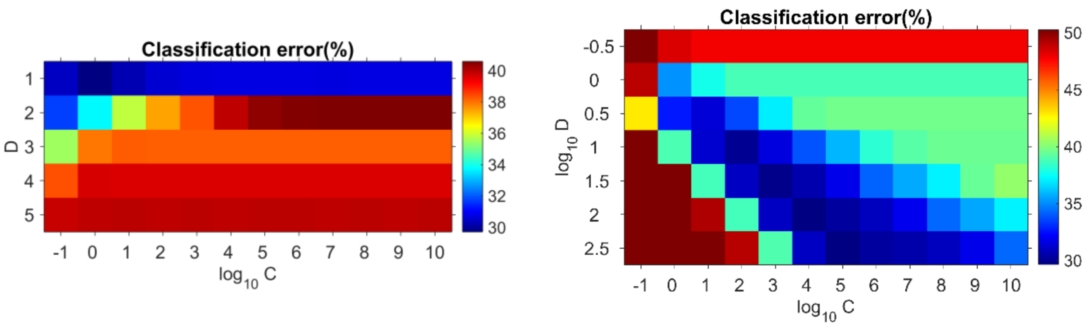

Figure 8: Average error rate of the classifier to predict the trend of the highest exchange rate from t-1 to t for daily data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

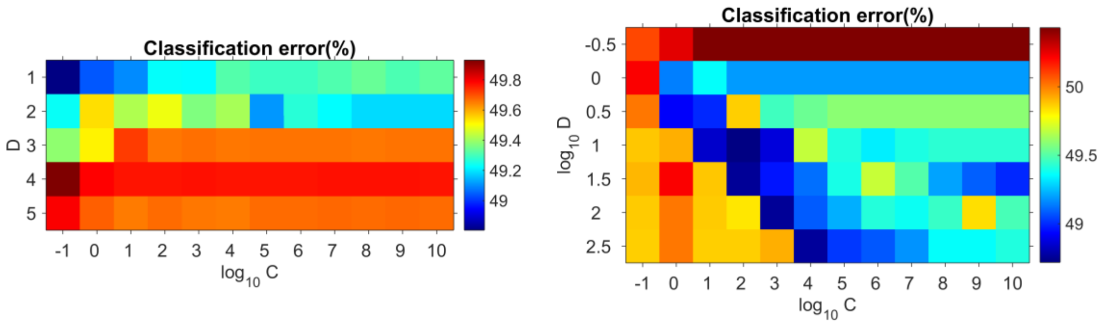

Figure 9: Average error rate of the classifier to predict the trend of the lowest exchange rate from t-1 to t for daily data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

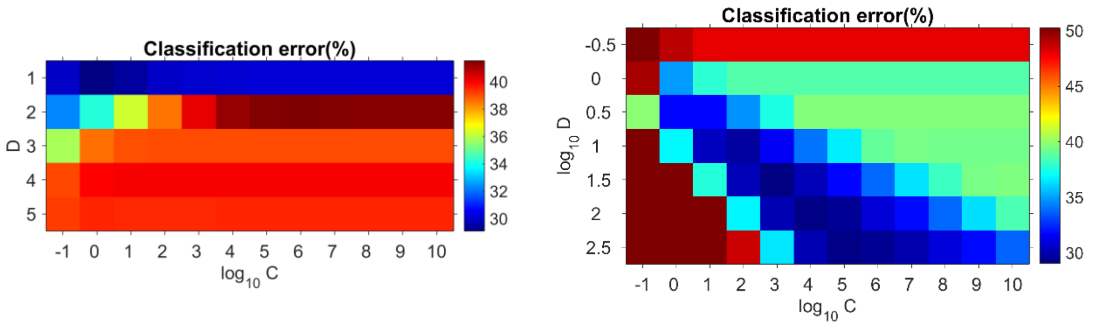

Figure 10: Average error rate of the classifier to predict the trend of the closing exchange rate from t-1 to t for daily data. Error rate is estimated using 10-fold cross-validation for classifier with different penalties for constraint violation C, and kernel parameters D. D represents the degree of the polynomial for polynomial kernel (left), and represents the standard deviation of the Gaussian function for Gaussian kernel (right).

SUMMARY

1. The results of this project demonstrates that SVM can be used to predict the trend of change of future highest and lowest currency exchange prices with decent performance (~35% error rate). Although one needs to take into account the transaction cost to determine whether this approach can really be used for currency trading to make profits.
2. In this project I only predict the trend of the change of the price without considering the magnitude of changes, which is also crucial for profitable trading. One

approach to incorporate the change magnitude information would be convert the problem into a multi-class classification problem. For example, categorize trend of change together with the magnitude of changes into several classes, and classify the data using multi-class SVM.

3. In this project I only use the historical price information without any external information. Including the external information from news as features might further improve the classification accuracy.

REFERENCES

1. 6.867 lecture notes.
2. Christopher Stone and Larry Bull. Foreign Exchange Trading using a Learning Classifier System. Technical report, University of the West of England Bristol, BS16 1QY, United Kingdom, 2004.
3. Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003.