Support Vector Machine Regression for Volatile Stock Market Prediction

Haiqin Yang, Laiwan Chan, and Irwin King

Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T. Hong Kong
{hqyang,lwchan,king}@cse.cuhk.edu.hk

Abstract. Recently, Support Vector Regression (SVR) has been introduced to solve regression and prediction problems. In this paper, we apply SVR to financial prediction tasks. In particular, the financial data are usually noisy and the associated risk is time-varying. Therefore, our SVR model is an extension of the standard SVR which incorporates margins adaptation. By varying the margins of the SVR, we could reflect the change in volatility of the financial data. Furthermore, we have analyzed the effect of asymmetrical margins so as to allow for the reduction of the downside risk. Our experimental results show that the use of standard deviation to calculate a variable margin gives a good predictive result in the prediction of Hang Seng Index.

1 Introduction

Support Vector Machine (SVM), based on Statistical Learning Theory, was first developed by Vapnik [4,6]. It has become a hot topic of intensive study due to its successful application in classification tasks [7,8] and regression tasks [5,3], specially on time series prediction [1] and financial related applications [2].

When using SVM in regression tasks, the Support Vector Regressor must use a cost function to measure the empirical risk in order to minimize the regression error. Although there are many choices of the loss functions to calculate the cost, e.g., least modulus loss function, quadratic loss function, etc., the ε-insensitive loss function is such a function that exhibits the sparsity of the solution [4]. Typically, this ε-insensitive loss function contains a fixed and symmetrical margin(FASM)term. When the margin is zero or very small, one runs into the risk of overfitting the data with poor generalization while when the margin is large, one obtains a better generalization at the risk of having higher testing error. For financial data, due to the embedded noise, one must set a suitable margin in order to obtain a good prediction. This paper focuses on two ways to set the margins in SVR.

When applying SVR to time series prediction, the practitioners usually overlook the choices of the margin setting. For example, in [2], they simply set the margin to 0. This amounts to the least modulus loss function. Others have just set the margin to a very small value [5,9,10]. In [1], they applied additional calculations, e.g., validation techniques, to determine a suitable margin empirically.

H. Yin et al. (Eds.): IDEAL 2002, LNCS 2412, pp. 391–396, 2002.

One of the shortcomings of the above methods is that the margin is symmetrical and fixed. Consequently, this technique is insensitive and non-adaptive to the input data. This may result in less-than-optimal performance in the testing data while it obtains a good result on the training data.

In this paper, we propose to use an adaptive margin in SVR for financial prediction to minimize the downside risk, which is an essential part in financial prediction with volatile financial data. More specifically, we present two approaches: one uses the fixed and asymmetrical margins(FAAM), whereas the other uses non-fixed and symmetrical margins(NASM).

A key difference between FAAM and FASM is that there exist an up and a down margin that are asymmetrical. In the case of FAAM when the up margin is greater than the down margin, the predictive results tend to be smaller than the predictive results which are produced by using FASM.

In NASM, the margin is adaptive to the input data. There are many possible choices to set the margin. For example, one may use the n-th order statistics to calculate the margin. More specifically, we choose the second order statistics, the standard deviation, as our method to calculate the adaptive margin. This is because that the standard deviation is frequently used as a measure of the volatility of stock prices in financial data. When the stock price is highly volatile, it has a high standard deviation. In financial time series the noise is often very large, and we try to tolerate our prediction by having a larger margin when the stock price is highly volatile. On the other hand, a smaller margin may be more suitable for less volatile stock activities. Hence, our approach avoids the fixed margin in order to obtain a better prediction result.

The paper is organized as follows. We introduce a general type of ε-insensitive loss function and give the inferential result in Section 2. We report experiments and results in Section 3. Lastly, we conclude the paper with a brief discussion and final remarks in Section 4.

2 Support Vector Regression

Given a training data set, $(x_1, y_1), \ldots, (x_N, y_N)$, where $x_i \in X, y_i \in R$, N is the size of training data, and X denotes the space of the input samples–for instance, R^n. The aim is to find a function which can estimate all these data well. SVR is one of the methods to perform the above regression task [4,3].

In general, the estimation function in SVR takes the following form,

$$f(x) = (w \cdot \phi(x)) + b, \tag{1}$$

where $(\cdot)$ denotes the inner product in Ω, a feature space of possibly different dimensionality such that $\phi : X \to \Omega$ and $b \in R$.

Now the question is to determine w and b from the training data by minimizing the regression risk, $R_{reg}(f)$, based on the empirical risk,

$$R_{reg}(f) = C \sum_{i=1}^{N} \Gamma(f(x_i) - y_i) + \frac{1}{2}(w \cdot w), \tag{2}$$

where C is a pre-specified value, $\Gamma(\cdot)$ is a cost function that measures the empirical risk. In general, the ε-insensitive loss function is used as the cost function [4]. For this function, when the data points are in the range of $\pm\varepsilon$, they do not contribute to the output error. The function is defined as,

$$\Gamma(f(x)-y)=\begin{cases} 0, & \text{if} \quad |y-f(x)|<\varepsilon \\ |y-f(x)|-\varepsilon, & \text{otherwise} \end{cases}. \tag{3}$$

In this paper, we introduce a general type of ε-insensitive loss function, which is given as,

$$\Gamma'(f(x_i)-y_i)=\begin{cases} 0, & \text{if } -\varepsilon_i^{down}<y_i-f(x_i)<\varepsilon_i^{up} \\ y_i-f(x_i)-\varepsilon_i^{up}, & \text{if } y_i-f(x_i)\geq\varepsilon_i^{up} \\ f(x_i)-y_i-\varepsilon_i^{down}, & \text{if } f(x_i)-y_i\geq\varepsilon_i^{down} \end{cases}, \tag{4}$$

where ε_i^{up} and ε_i^{down} correspond to the i-th up margin and down margin respectively. When ε_i^{up} and ε_i^{down} are both equal to a constant, for all $i, i=1,\ldots,N$, Eq. (4) amounts to the ε-insensitive loss function in Eq. (3) and it is labeled as FASM (Fixed and Symmetrical Margin). When $\varepsilon_i^{up}=\varepsilon^{up}$, for all $i=1,\ldots,N$ and $\varepsilon_j^{down}=\varepsilon^{down}$, for all $j=1,\ldots,N$ with $\varepsilon^{up}\neq\varepsilon^{down}$, this case is labeled as FAAM (Fixed and Asymmetrical Margin). In the case of NASM (Non-fixed and Symmetrical Margin), we use an adaptive margin for which the up margin equals to the down margin. The last case is with an adaptive and asymmetrical margin. In this paper, we just consider the first three cases, i.e., FASM, FAAM, and NASM.

Using the Lagrange function method to find the solution which minimizes the regression risk of Eq. (2) with the cost function in Eq. (4), we obtain the following Quadratic Programming (QP) problem:

$$\arg\min_{\alpha,\alpha^*}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(\phi(x_i)\cdot\phi(x_j))+\sum_{i=1}^{N}(\varepsilon_i^{up}-y_i)\alpha_i + \sum_{i=1}^{N}(\varepsilon_i^{down}+y_i)\alpha_i^* \tag{5}$$

subject to

$$\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)=0, \alpha_i, \alpha_i^*\in[0,C], \tag{6}$$

where α and α^* are corresponding Lagrange multipliers used to push and pull $f(x_i)$ towards the outcome of y_i respectively.

Solving the above QP problem of Eq. (5) with constraints of Eq. (6), we determine the Lagrange multipliers α and α^* and obtain $w=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)\phi(x_i)$. Therefore the estimation function in Eq. (1) becomes

$$f(x)=\sum_{i=1}^{N}(\alpha_i-\alpha_i^*)(\phi(x)\cdot\phi(x_i))+b. \tag{7}$$

So far, we have not considered the computation of b. In fact, this can be solved by exploiting the Karush-Kuhn-Tucker(KKT) conditions. These conditions state that at the optimal solution, the product between the Lagrange multipliers and the constraints has to equal to zero. In this case, it means that

$$\begin{aligned} \alpha_i(\varepsilon_i^{up} + \xi_i - y_i + (w \cdot \phi(x_i)) + b) = 0 \\ \alpha_i^*(\varepsilon_i^{down} + \xi_i^* + y_i - (w \cdot \phi(x_i)) - b) = 0 \end{aligned} \tag{8}$$

and

$$\begin{aligned} (C - \alpha_i)\xi_i = 0 \\ (C - \alpha_i^*)\xi_i^* = 0. \end{aligned}$$

where ξ_i and ξ_i^* are slack variables used to measure the error of up side and down side.

Since $\alpha_i \cdot \alpha_i^* = 0$ and $\xi_i^{(*)} = 0$ for $\alpha_i^{(*)} \in (0, C)$, b can be computed as follows:

$$b = \begin{cases} y_i - (w \cdot \phi(x_i)) - \varepsilon_i^{up}, & \text{for } \alpha_i \in (0, C) \\ y_i - (w \cdot \phi(x_i)) + \varepsilon_i^{down}, & \text{for } \alpha_i^* \in (0, C) \end{cases}. \tag{9}$$

Using the trick of kernel function, Eq. (7) can be written as, $f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x, x_i) + b$, where the kernel function, $K(x, x_i) = (\phi(x) \cdot \phi(x_i))$, which is a symmetric function and satisfies the Mercer's condition. In this paper, we select a common kernel function, e.g., RBF function, $K(x, x_i) = \exp(-\beta|x - x_i|^2)$, as the kernel function.

In the next section, we apply our inferential result of SVR based on the general type of ε-insensitive loss function to the regression of financial data, for example, indices and stock prices. By applying regression to the data, we can build a dynamic system to model the data and hence use the system for predicting future prices.

3 Experiments

In this section, we conduct two experiments to illustrate the effect of FASM, FAAM, and NASM. The first experiment illustrates the SVM financial prediction with fixed margin, including FASM and FAAM. The second experiment tests the SVM financial prediction with NASM under shift windows.

In our experiment, we use the daily closing price of Hong Kong's Hang Seng Index (HSI) from January 15, 2001 to June 19, 2001, a total of 104 days' of data points, out of which 100 data points for training and testing. We set the length of the shift window to 80. The dynamic system is modeled as $\widehat{I_t} = f(I_{t-4}, I_{t-3}, I_{t-2}, I_{t-1})$, where I_t is the real stock price at time t, and $\widehat{I_t}$ is the predictive value at time t. Therefore, the first training data set is from January 15, 2001 to May 22, 2001, a total of 84 days' of HSI. We use them to predict the next day's HSI. This window is then shifted and an entire training is performed again to predict the following day's HSI for the remaining testing data.

The SVR algorithm used in our experiment is modified from LibSVM [10]. Before running the algorithm, we need to determine some parameters. They are C, the cost of error; β, parameter of kernel function, and the margins. After performing a cross-validation in the first training data, we set $C = 6000$, $\beta = 2^{-24}$. Since different margins will affect the results of prediction, we use different values in our tests. Furthermore, we use the following three error definitions to measure the testing errors, $error \equiv \frac{1}{M}\sum_{t=1}^{M}|I_t - \widehat{I}_t|, error_{pos} \equiv \frac{1}{M}\sum_{t=1, I_t \geq \widehat{I}_t}^{M}(I_t - \widehat{I}_t), error_{neg} \equiv \frac{1}{M}\sum_{t=1, I_t < \widehat{I}_t}^{M}(\widehat{I}_t - I_t)$, where M is the size of the testing data and $error$ reflects the total risk, $error_{pos}$ reflects the upside risk and $error_{neg}$ reflects the downside risk respectively.

The experiments are conducted on a Pentium 4, with 1.4 GHZ, 512M RAM and Windows2000. With these configurations, the predictive results are obtained within seconds.

In the first experiment we use different values for up margin and down margin to test the effect of FASM and FAAM. We show the setting of the margin in the second and third columns of Table 1, and report the corresponding errors in the last three columns. In all but the first and the last margin setting, their overall margin widths are the same, i.e., $\varepsilon^{up} + \varepsilon^{down} = 150$. This allows us to have a fair comparison of the four cases. From the Table 1, we can see that the $error_{pos}$ gradually increases with the increase of ε^{up}. At the same time, with the increase of ε^{up}, we allow for more errors above the predictive values. Thus the $error_{neg}$ decreases. In terms of the overall error, it increases and then decreases again. This indicates that neither a narrow margin for the upside nor the downside would be desirable in terms of the overall error.

In the second experiment, after considering the volatility of the financial data, we set the up margin and down margin both equal to the standard deviation of the input vector x to perform the prediction. The predictive error of the experiment with the NASM is reported in the last row of Table 1 and the result shows that the total error is significantly decreased comparing with the fixed ones.

Table 1. Experiment Results

Case	ε^{up}	ε^{down}	$error$	$error_{pos}$	$error_{neg}$
1	0	0	134.59	56.46	78.13
2	50	100	131.96	49.44	82.52
3	75	75	129.03	60.47	68.56
4	100	50	129.96	73.44	56.52
5	150	0	135.64	101.28	34.36
6	σ	σ	**116.19**	53.29	62.90

4 Discussion and Conclusion

In this paper, we present a general type of ε-insensitive loss function in SVR and outline the various margins used, i.e., FASM, FAAM and NASM. Using Hong Kong's HSI as the data set for SVR with different types of margins, we have the following conclusions:

1. One interesting observation is that neither the up margin nor the down margin would affect the $error$ unilaterally. This can be seen from the results of Case 2 to Case 5 in Table 1.
2. Another interesting observation is that from the point of view of the downside risk, Case 5 in Table 1 is a good result since its $error_{neg}$, which is related to the downside risk, is minimum. In practice, we can reduce the downside risk by increasing the up margin while decreasing the down margin.
3. In the NASM case, we find that using standard deviation to calculate the margin, which can reflect the change in volatility of the financial data, results in the best prediction in our experiment since this result has a minimal $error$.

Acknowledgement. The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administration Region, China. The authors thank Lin, Chih-Jen for helpful suggestions on using the LibSVM.

References

1. K. R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen and V. N. Vapnik. Predicting time series with support vector machines. ICANN, 999-1004, 1997.
2. T. B. Trafalis and H. Ince. Support vector machine for regression and applications to financial forecasting. IJCNN2000, 348-353.
3. A. Smola and B. Schölkopf. A Tutorial on Support Vector Regression. 1998, Technical Report NeuroCOLT NC-TR-98-030.
4. V. N. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, 1995.
5. V. N. Vapnik, S. Golowich and A. Smola. Support vector method for function approximation, regression estimation and signal processing.
6. V. N. Vapnik. Statistical Learning Theory. Wiley, New York, 1998.
7. Edgar Osuna and Robert Freund and Federico Girosi. Support Vector Machines: Training and Applications. AIM-1602, MIT, 38, 1997.
8. Christopher J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2):121-167, 1998.
9. S. Mukherjee, E. Osuna and F. Girosi. Nonlinear prediction of chaotic time series using support vector machines. IEEE Workshop on Neural Networks for Signal Processing VII, IEEE Press, J. Principe and L. Giles and N. Morgan and E. Wilson, 511, 1997.
10. Chih-Chung, Chang and Chih-Jen, Lin. LIBSVM: a Library for Support Vector Machines (Version 2.31), 2001.
11. Frank A. Sortino and Stephen E. Satchell. Managing downside risk in financial markets : theory, practice, and implementation. Oxford, Boston:Butterworth-Heinemann, 2001.