'Enforced-Denial Support-Vector Machines for Noisy Data With Applications to Financial Time Series Forecasting

David Edelman
University of Wollongong

February 7, 2002

Abstract

The Support-Vector Machine (SVM) approach of Vapnik has received much attention recently as being a simple and effective approach for pattern recognition and classification in problems with low-to-moderate noise component. While the strength of SVM's lies in extracting definition in 'hard-to-classify' cases at the 'boundaries' between categories, it has been noted that their error-penalty structure, which sacrifices modelling suitability in favour of computational tractability, is not ideal, and indeed tends to lead to inferior models in problems of high noise component. In the present paper, we explore a method for improving the error-penalty character of SVM's in the presence of noisy data, while preserving the computational tractability of the SVM approach. The method is demonstrated on a Time Series dataset of the Australian All-Ordinaries Index, with the results then compared to those of other traditional and nonlinear methods applied to the same dataset.

1 Introduction

Since their invention by Vapnik in 1979 (see Vapnik[6][7]) Support Vector Machines have been found to perform well in pattern recognition problems such as face[2] and character[3] recognition. As SVM methods are based on locating boundary cases between regions such as classification regions, the resulting models have low complexity, these 'borderline' cases typically being few in number relative to the total size of datasets used for training.

While the initial formulation was based on perfectly classifiable or 'separable' groups, an extension for problems of moderate to low proportion of misclassification or 'noise' was developed. The cost function introduced for errors was of the form of a sum of distances to the boundary of 'correct' classification region. While the modeling weakness of this error cost function were conceded, its qualitative adequacy and computational convenience were deemed to be sufficient justification for its adoption, with the proviso that the rate of errors in the underlying problem be 'small'. A wide range of anecdotal evidence has indeed shown that for problems of large error rate (such as Financial Forecasting) the misclassification error criterion causes SVM's to behave particularly poorly in relation to other methods.

Possibly for this reason, researchers in Financial Forecasting have been slow to adopt Support Vector Machine methods, though a few studies involving applications to the Financial Markets are beginning to appear, such as Tay and Cao (2001)[5], who found SVM's superior to Backpropagation Neural Network models in forecasting several Chicago-based Futures prices, including those of the S&P500 index.

In what follows, an approach will be presented which addresses the poor performance of SVM's for high-noise problems, by extending the intuition which originally resulted in the SVM approach; just as 'easy to classify cases' were seen to be irrelevant to decision boundaries in seperable problems, 'impossible to classify' cases (or violations of seperability) should not affect the location of decision boundaries in the non-seperable case.

2 Background

Before presenting the new results of the present paper, a brief summary and introduction to the concepts and notation of the classic SVM problem will be mentioned.

Briefly, of the various steps involved in applying Support-Vector Machines, the first involves (uncharacteristically, for statistical modeling) expanding the input space to what is referred to as *feature* space: for a given input vector $\underline{x}$,

$$\underline{x} \to \underline{\phi}(\underline{x}),$$

typically, as here, with Radial Basis Functions centered around each input training example:

$$\phi_j(\underline{x}_i) = \exp\{-\frac{1}{2}||\underline{x}_i - \underline{x}_i||^2\},\ j = 1, 2, \ldots, n$$

Next, a discriminant for deciding whether $y = 1$ or $y = -1$ is defined by

$$\underline{w}'\underline{\phi} + b$$

where for a sample $(\underline{x}_1, y_1), \ldots, (\underline{x}_n, y_n)$ $\underline{w}$ is chosen such that $\frac{1}{2}\underline{w}'\underline{w} + C\sum \xi_i$ is minimised subject to

$$y_i(\underline{w}'\underline{\phi}_i + b) \geq 1 - \xi_i,\ i = 1, 2, \ldots, n$$

$\xi_i \geq 0$, $i = 1, 2 \ldots, n$, and C is a suitably chosen 'slack' constant.

This is a well-posed Quadratic Programming (QP) problem, most easily solved by solving a 'dual' QP problem, with the net result that a small proportion of input vectors define the 'Separating Hyperplane', or Discriminant function. The value of the 'slack constant' C is usually adjusted by monitoring the generalisation performance on a holdout sample.

The method proposed here involves solving a sequence of problems of exactly the above form, but where in each case, the succeeding problem to be solved is identical, except for the removal of the data example corresponding to the largest of the resulting ξ's. This procedure is repeated until no nonzero ξ's remain. While this increases computing time by a considerable degree, it is nevertheless tractable, and, in small problems investigated anecdotally, seems likely to lead to solutions similar to the (non-QP) variant of minimising $\frac{1}{2}\underline{w}'\underline{w} + C\sum H(\xi_i)$ is minimised subject to

$$y_i(\underline{w}'\underline{\phi}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,$$

where $H(t)$ is the Heaviside function, equal to 0 for $t \leq 0$, and 1 for $t > 0$.

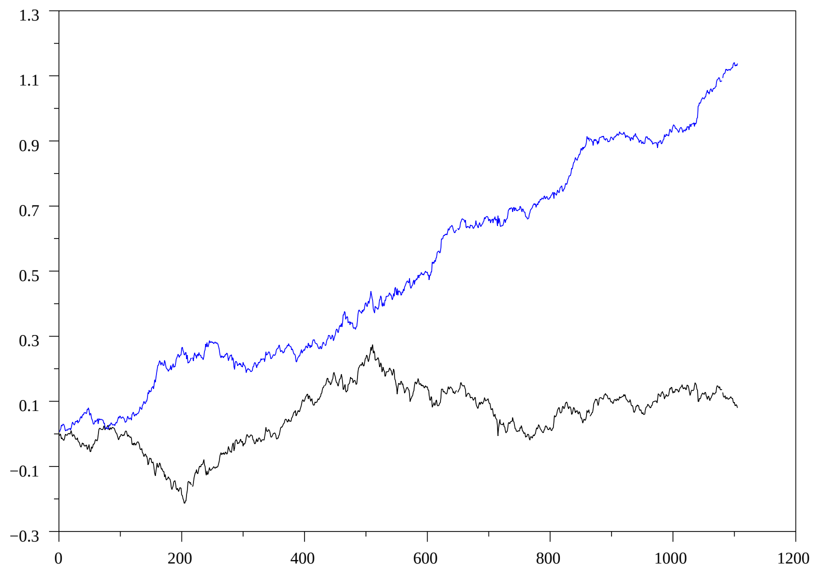
For the present dataset, consisting of 17 partially overlapping time windows of 260-day history of the All-Ordinaries Index, each used to forecast the 65 days succeeding, the method was applied, where in each case, the inputs consisted of the returns for the 10 previous trading days.

Using the numerical package SCILAB[4] on a 1.2Ghz Linux machine with 512MB memory, the forecasts were obtained in approximately 2 minutes per window.

The cumulative returns for the period are graphed, along with the cumulative returns for the 'Buy and Hold' strategy, in Figure 2.

3 Discussion

While the results presented here are preliminary, it appears that this extension of the SVM technique is successful in forecasting returns for this dataset, a significant improvement on the results obtained using standard SVM methods (found to be so poor that they are not presented here). Also, while the emphasis here was to demonstrate the new technique for applying SVM's to particularly noisy data, it is perhaps worth noting that the excess returns achieved above in relation to the 'Buy and Hold' strategy, are significant enough not to be eliminated when the effects of reasonable transaction costs (including bid-offer slip) are included, suggesting the presence of a Weak Form inefficiency.



It is hoped that this new technique will help increase the applicability of Support Vector Machines to 'high' noise problems in future.

References

[1] Chung, V. (1999). "Using Neural Networks to Predict the Returns of the Australian All Ordinaries Index", *Honours' Thesis, Department of Accounting and Finance, University of Wollongong*

[2] Guo, G., S.Z. li, K.L. Chan (2001), "Support Vector Machines for Face Recognition", *Image and Vision Computing*, Vol 19, pp.631-638

[3] Joachims T. (1997), "Text and Characterisation with Support Vector Machines", *LS VIII Technical Rept.* no. 23, University of Dorismund

[4] *SCILAB: a numerical computing package from INRIA* http://www.inria.fr

[5] Tay, E.H. and L.J. Cao (2001), "Application of Support Vector Machines in Financial Time Series Forecasting", *Omega 29*, pp.309-317

[6] Vapnik, V.N. (1995), *The Nature of Statistical Learning Theory* New York: Springer

[7] Vapnik, V.N. (1995), *Statistical Learning Theory* New York: Wiley Inter-science

[8] Walmsley (1999). "Using Neural Networks to Arbitrage the Australian All Ordinaries Index", *Honours' Thesis, Department of Accounting and Finance, University of Wollongong*