Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 36 (2009) 132–138
Expert Systems with Applications
www.elsevier.com/locate/eswa

Regularized least squares fuzzy support vector regression for financial time series forecasting

Reshma Khemchandani [a], Jayadeva [b], Suresh Chandra [a,*]

[a] Department of Mathematics, Indian Institute of Technology, Hauz Khas, New Delhi 110 016, India
[b] IBM India Research Lab, Block-C, Institutional Area Vasant Kunj, New Delhi 110 070, India

Abstract

In this paper, we propose a novel approach, termed as regularized least squares fuzzy support vector regression, to handle financial time series forecasting. Two key problems in financial time series forecasting are noise and non-stationarity. Here, we assign a higher membership value to data samples that contain more relevant information, where relevance is related to recency in time. The approach requires only a single matrix inversion. For the linear case, the matrix order depends only on the dimension in which the data samples lie, and is independent of the number of samples. The efficacy of the proposed algorithm is demonstrated on financial datasets available in the public domain.

Keywords: Machine learning; Support vector machines; Regression; Financial time series forecasting; Fuzzy membership

1. Introduction

Over the last decade, support vector machines (SVMs) have emerged as the paradigm of choice for pattern classification and regression (Burges, 1998; Cristianini & Shawe-Taylor, 2000; Smola, 1996; Vapnik, 1998). SVMs emerged from research in statistical learning theory on how to trade off structural complexity against empirical risk. The "maximum margin SVM" is amongst the most popular of SVM classifiers. It aims to minimize an upper bound on the generalization error through maximizing the margin between two disjoint half planes (Burges, 1998; Vapnik, 1998).

Support vector regression (SVR) aims to fit a linear regressor through a given set of data points, where the points may be in the pattern space or in a higher dimensional feature space. Determining such a regressor requires solving a quadratic minimization problem subject to linear inequality constraints, which is a convex programming task (Burges, 1998).

Typical technical data involved in financial time series prediction are close value (price of the last performed trade during the day), highest traded price during the day, lowest traded price during the day, and volume (total number of traded stock during the day) (Pissarenko, 2002). Financial time series illustrate regime shifting, i.e. their statistical properties vary with time (the process is time-varying (Hellström & Holmström, 1998; Pissarenko, 2002)).

A common property of financial time series is volatility clustering, i.e. large changes tend to succeed large ones, while small changes are followed by small ones. This, combined with a lack of long-term stationarity, suggests that a conventional SVM approach that lays equal emphasis on samples in a sequence would find it does hard to capture any input–output relationship inherent in the data. It also indicates that any attempt at predicting a future sample from past ones might benefit by giving more emphasis to recent samples, as against older ones. Some fuzzy SVM techniques have attempted to overcome this problem (e.g. Jayadeva, Khemchandani, & Chandra, 2004; Lin & Wang, 2002; Lin & Wang, 2005).

* Corresponding author. Tel.: +91 11 26591479; fax: +91 11 26581005.
E-mail addresses: reshmaiitd@gmail.com (R. Khemchandani), jayadeva@ee.iitd.ac.in (Jayadeva), chandras@maths.iitd.ac.in (S. Chandra).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.09.035

Taking motivation from Jayadeva et al. (2004), Lin and Wang (2002), Lin and Wang (2005), Tay and Cao (2002), we propose a novel approach to support vector regression for financial time series forecasting, whose preliminary ideas have been reported in Jayadeva, Khemchandani, and Chandra (2006). This approach is motivated by the observation that in a non-stationary time series, the dependency between the input variables and the output does not remain constant over time. Specifically, recent data samples provide more relevant information than more distant ones. The proposed approach assigns fuzzy membership values to data samples. This not only helps capture the input–output relationship in a better way, but also reduces overfitting. The proposed approach requires only one matrix inversion to determine the regressor. Experimental results on selected financial datasets available in the public domain show that the proposed approach not only yields comparable results, but is also faster than other reported approaches.

The paper is organized as follows: Section 2 briefly dwells on support vector regression and also introduces the notation that is employed in the rest of the paper. Section 3 discusses least squares support vector machines. Section 4 proposes regularized least squares fuzzy support vector regression, and contains some theoretical results. Section 5 is devoted to experimental results. Section 6 is devoted to concluding remarks.

2. Support vector regression

We first discuss support vector regression, prior to introducing our approach. We next highlight the differences between our approach and conventional support vector regression.

Consider a set of M data samples, and let the ith sample be represented by the tuple (P_i, y_i), where $P_i = (P_{i1}, P_{i2}, \ldots, P_{iN})$ is a point in $\mathbf{R}^N$ and y_i is the value at point P_i. Let P denote the matrix comprised of the row vectors P_i, $i = 1, 2, \ldots, M$. The goal of regression is to find a functional relationship $y = h(x)$ between $x \in \mathbf{R}^N$ and $y \in \mathbf{R}$. The regressor must not only fit the given data well, but also make minimal errors in predicting the values at any other arbitrary point in $\mathbf{R}^N$.

Nonlinear regression is accomplished by fitting a linear regressor in a higher dimensional feature space. A nonlinear transformation ϕ is used to transform data points from the input space of dimension N into a feature space having a higher dimension L. The nonlinear mapping is denoted by ϕ: $\mathbf{R}^N \to \mathbf{R}^L$, so that $\phi(P_i) \in \mathbf{R}^L$, $i = 1, 2, \ldots, M$.

Conventional SVR solves the optimization problem

$$\text{(SVR)} \quad \text{Minimize}_{q_1,q_2,w,b,} \quad C(e^{\mathrm{T}}q_1 + e^{\mathrm{T}}q_2) + \frac{1}{2}w^{\mathrm{T}}w \tag{1}$$

$$\begin{aligned} \text{subject to} \quad & w^{\mathrm{T}}\phi(P_i) + b - y_i \leqslant \epsilon + q_{1i}, \\ & i = 1, 2, \ldots, M, \\ & y_i - w^{\mathrm{T}}\phi(P_i) - b \leqslant \epsilon + q_{2i}, \\ & i = 1, 2, \ldots, M, \\ & q_{1i}, q_{2i} \geqslant 0, \quad i = 1, 2, \ldots, M, \end{aligned} \tag{2}$$

where $C > 0$ is a parameter. The solution to (1) and (2) is determined by solving its dual, which is given by

$$\text{Maximize}_{\beta,\alpha} \quad y^{\mathrm{T}}(\beta - \alpha) - (\beta - \alpha)^{\mathrm{T}}PP^{\mathrm{T}}(\beta - \alpha) - \epsilon e^{\mathrm{T}}(\beta + \alpha) \tag{3}$$

$$\begin{aligned} \text{subject to} \quad & e^{\mathrm{T}}(\beta - \alpha) = 0, \\ & 0 \leqslant \beta, \quad \alpha \leqslant C. \end{aligned} \tag{4}$$

Note that the data samples P appear in inner product form and therefore, the problem (3) and (4) can be rewritten as

$$\text{Maximize}_{\beta,\alpha} \quad y^{\mathrm{T}}(\beta - \alpha) - (\beta - \alpha)^{\mathrm{T}}K(\beta - \alpha) - \epsilon e^{\mathrm{T}}(\beta + \alpha) \tag{5}$$

$$\begin{aligned} \text{subject to} \quad & e^{\mathrm{T}}(\beta - \alpha) = 0, \\ & 0 \leqslant \beta, \quad \alpha \leqslant C, \end{aligned} \tag{6}$$

where the matrix K is termed as a kernel matrix and its elements are given by $(K_{ij}) = [\phi(P_i)]^{\mathrm{T}}\phi(P_j)$, $i,j = 1, 2, \ldots, M$.

3. Least squares support vector machines

Suykens and Vandewalle (1999) introduced least squares support vector machines (LS-SVMs), whose formulation employs equality type constraints. This allows the solution to be found by solving a set of linear equations, instead of the quadratic programming problem that classical SVMs solve.

LSSVM regression involves solving the following optimization problem

$$\text{(LSSVM)} \quad \text{Minimize}_{q,w,b} \quad \frac{C}{2}\|q\|^2 + \frac{1}{2}w^{\mathrm{T}}w, \tag{7}$$

$$\begin{aligned} \text{subject to} \quad & w^{\mathrm{T}}\phi(P_i) + b - y_i + q_i = 0, \\ & i = 1, 2, \ldots, M. \end{aligned} \tag{8}$$

where $C > 0$ is a parameter. The Lagrangian (Mangasarian, 1994) for the problem (7) and (8) is given by

$$L(w, b, q, \alpha) = \frac{1}{2}(w^{\mathrm{T}}w) + \frac{C}{2}\|q\|^2 - \sum_{i=1}^{M}\alpha_i(w^{\mathrm{T}}\phi(P_i) + b + q_i - y_i), \tag{9}$$

where α_i's are the Lagrangian multipliers that may be of any sign, as follows from the equality constraints of the Karush–Kuhn–Tucker (K–K–T) conditions. Simplifying, these conditions may be compactly written in the form

$$\begin{bmatrix} 0 & e^{\mathrm{T}} \\ e & K + \frac{I}{C} \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}, \tag{10}$$

where $Y = [y_1, y_2, \ldots, y_M]$; e is a vector of ones of appropriate dimension, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_M]$ are the Lagrange parameters, I is an identity matrix of appropriate dimension and K denotes the kernel matrix. The linear kernel, with

$$K(P_i, P_j) = \langle P_i, P_j \rangle, \tag{11}$$

the polynomial one, where

$$K(P_i, P_j) = (1 + \langle P_i, P_j \rangle)^2, \tag{12}$$