Expert Systems with Applications 37 (2010) 8784–8792

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

On Lagrangian support vector regression

S. Balasundaram *, Kapil

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

ARTICLE INFO

Keywords:
Lagrangian support vector machines
Support vector regression
Time series

ABSTRACT

Prediction by regression is an important method of solution for forecasting. In this paper an iterative Lagrangian support vector machine algorithm for regression problems has been proposed. The method has the advantage that its solution is obtained by taking the inverse of a matrix of order equals to the number of input samples at the beginning of the iteration rather than solving a quadratic optimization problem. The algorithm converges from any starting point and does not need any optimization packages. Numerical experiments have been performed on Bodyfat and a number of important time series datasets of interest. The results obtained are in close agreement with the exact solution of the problems considered clearly demonstrates the effectiveness of the proposed method.

1. Introduction

Support vector machine (SVM) methods based on statistical learning theory (Vapnik, 2000) have been successfully applied to many problems of practical importance (Guyon, Weston, Barnhill, & Vapnik, 2002; Osuna, Freund, & Girosi, 1997) due to its high generalization performance over other learning methods. It is well known that the standard SVM formulation (Burges, 1998; Cristianini & Shawe-Taylor, 2000) leads to the solution of a quadratic programming problem with linear inequality constraints and that the problem will have an unique solution. With the combined advantages of generalization performance and unique solution SVM becomes an attractive method on problems of interest.

The goal of regression problem is in determining the underlying mathematical relationship between the given input observations and their output values. Regression models have been successfully applied in many important fields of study such as economics, engineering and bioinformatics. By the introduction of ε-insensitive error loss function proposed by Vapnik (2000) SVM methods have been successfully applied to regression problems (Mukherjee, Osuna, & Girosi, 1997; Muller, Smola, Ratsch, Schölkopf, & Kohlmorgen, 1999; Tay & Cao, 2001).

Considering the 2-norm error loss function instead of the usual 1-norm and maximizing the margin with respect to both the orientation and the relative location to the origin of the bounding planes, Mangasarian and Musicant (2001a, 2001b) studied "equivalent" SVM formulations for classification problems. This formulation leads to solving a positive-definite dual problem having only the non-negative constraints of the dual variables. Further their work on the study of formulating the problem of machine learning and data mining as an unconstrained minimization problem whose objective function is strongly convex and obtaining its solution using finite Newton method (see Fung & Mangasarian, 2003; Mangasarian, 2002). Since the objective function is not twice differentiable in this formulation by applying a smoothing technique a new SVM formulation called smooth SVM (SSVM) has been proposed in Lee and Mangasarian (2001). For the study on an extension of SSVM to ε-insensitive error loss based support vector regression (SVR) problems (see Lee, Hsieh, & Huang, 2005). Finally for the extension of the Active set SVM (ASVM) (Mangasarian & Musicant, 2001b) method proposed for classification problems to SVR problems we refer the reader to Musicant and Feinberg (2004).

Motivated by the study of Lagrangian SVM (Mangasarian & Musicant, 2001a) for classification problems we propose in this paper Lagrangian ε-insensitive SVR formulation. The main advantage of our approach in comparison with the standard SVR formulation is that the solution of the problem is obtained by taking the inverse of a matrix at the beginning of the iteration rather than solving a quadratic programming problem. In order to verify the effectiveness of the proposed method a number of problems of practical importance are considered. It is observed that the results obtained are in close agreement with the exact solution of the problems considered.

Throughout in this work all the vectors are assumed as column vectors. For any two vectors x, y in the n-dimensional real space R^n the inner product of the vectors will be denoted by $x^t y$ where x^t is the transpose of the vector x. When x is orthogonal to y we write $x \perp y$. The 2-norm of a vector x and a matrix Q will be denoted by $\|x\|$ and $\|Q\|$ respectively. For any vector $x \in R^n$, x_+ is a vector in R^n obtained by setting all the negative components of x to zero. For matrices $M \in R^{m\times n}$ and $N \in R^{n\times \ell}$, the kernel matrix K of size

* Corresponding author. Tel.: +91 11 26704724; fax: +91 11 26741586.
E-mail addresses: balajnu@hotmail.com, balajnu@gmail.com (S. Balasundaram).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.06.028

$m \times \ell$ is denoted by $K = K(M,N)$. The identity matrix of appropriate size is denoted by I and the column vector of ones of dimension m by e.

The paper is organized as follows. In Section 2 the linear and nonlinear SVR formulations for the standard 1-norm and 2-norm are introduced. By considering the Karush–Kuhn–Tucker (KKT) conditions the Lagrangian SVR algorithm is formulated in Section 3 and its convergence follows from the result of Mangasarian and Musicant (2001a). Numerical experiments have been performed on Bodyfat, Mackey–Glass, IBM, Google, Citigroup datasets and their results are compared with the exact solutions in Section 4. Finally we conclude our work in Section 5.

2. Support vector regression formulation

In this section, we briefly describe the standard 1-norm and 2-norm SVR formulations. For the given set of input samples $\{(x_i,y_i)\}_{i=1,2,\ldots,m}$ where $x_i \in R^n$ and $y_i \in R$, the linear SVR problem is the method in approximating the output by a function $f(\cdot)$ of the form:

$$f(x) = x^t w + b, \tag{1}$$

where $w \in R^n$ and $b \in R$ are determined as the solution of the following constrained, quadratic optimization problem with parameters $\nu > 0$ and $\varepsilon > 0$, written in matrix form as:

$$\min_{(w,b,\xi,\xi^*)\in R^{n+1+m+m}} \frac{1}{2} w^t w + \nu(e^t \xi + e^t \xi^*)$$

subject to

$$y - Aw - be \leqslant \varepsilon e + \xi,$$
$$Aw + be - y \leqslant \varepsilon e + \xi^*$$

and

$$\xi_i, \xi_i^* \geqslant 0 \quad \text{for } i = 1,2,\ldots,m, \tag{2}$$

where $\xi = (\xi_1,\ldots,\xi_m)^t, \xi^* = (\xi_1^*,\ldots,\xi_m^*)^t$ are vectors of slack variables, $y = (y_1,\ldots,y_m)^t$ is the vector of observed values and $A \in R^{m\times n}$ be the matrix whose ith row, denoted by A_i, is defined to be the vector x_i^t. By introducing Lagrange multipliers $u_1, u_2 \in R^m$ the dual of the above problem (2) can be formulated as:

$$\min_{u_1,u_2\in R^m} \frac{1}{2}(u_1-u_2)^t AA^t (u_1-u_2) - y^t(u_1-u_2) + \varepsilon e^t(u_1+u_2)$$

subject to

$$e^t(u_1-u_2) = 0 \text{ and } 0 \leqslant u_1, u_2 \leqslant \nu. \tag{3}$$

Further, for any example $x \in R^n$ its prediction is given by the following function

$$f(x) = x^t A^t (u_1-u_2) + b.$$

For the nonlinear case, the input data is mapped into a higher dimensional feature space using a kernel function $k(\cdot,\cdot)$ and here in the feature space a linear SVR estimation is obtained. In this case, the support vector regression method will lead to the solution of the following Lagrangian problem:

$$\min_{(u_1,u_2)\in R^{m+m}} \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m (u_{1i}-u_{2i})k(x_i,x_j)(u_{1j}-u_{2j}) - \sum_{i=1}^m y_i(u_{1i}-u_{2i}) + \varepsilon \sum_{i=1}^m (u_{1i}+u_{2i})$$

subject to

$$e^t(u_1-u_2) = 0 \text{ and } 0 \leqslant u_1, u_2 \leqslant \nu, \tag{4}$$

where $k(\cdot,\cdot)$ is the kernel function and $u_1 = (u_{11},\ldots,u_{1m})^t$, $u_2 = (u_{21},\ldots,u_{2m})^t$ in R^m are the Lagrange multipliers. For example, for the Gaussian kernel

$$k(x_i,x_j) = \exp(-\mu\|x_i - x_j\|^2) \text{ for } i,j = 1,\ldots,m$$

and μ is a parameter.

Using the solution of the problem (4), the regression estimation function $f(\cdot)$ for the nonlinear case is obtained to be: For any input $x \in R^n$,

$$f(x) = \sum_{i=1}^m (u_{1i}-u_{2i})k(x,x_i) + b.$$

Following the approach of Mangasarian and Musicant (2001a, 2001b), Musicant and Feinberg (2004), by considering the square of the 2-norm of the slack variables ξ, ξ^* instead of 1-norm and adding the term $\left(\frac{b^2}{2}\right)$ in the definition of the objective function of (2) consider the linear SVR formulation with ε-insensitive error loss function (Vapnik, 2000) as a constrained minimization problem of the following form (Musicant & Feinberg, 2004):

$$\min_{(w,b,\xi,\xi^*)\in R^{n+1+m+m}} \frac{1}{2}(w^t w + b^2) + \frac{\nu}{2}\sum_{i=1}^m (\xi_i^2 + \xi_i^{*2})$$

subject to

$$\begin{aligned} &(y_i - A_i w - b) \leqslant (\varepsilon + \xi_i),\\ &(A_i w + b - y_i) \leqslant (\varepsilon + \xi_i^*) \quad i = 1,2,\ldots,m, \end{aligned} \tag{5}$$

where ξ_i, ξ_i^* are slack variables and ε, ν are input parameters. Since none of the components of the vector $\xi = (\xi_1,\ldots,\xi_m)^t$ or $\xi^* = (\xi_1^*,\ldots,\xi_m^*)^t$ will be negative at optimality (Musicant & Feinberg, 2004) their non-negativity constraints have been dropped in the formulation (5). Since the linear regression estimation function takes the form (1) its approximation to the vector $y \in R^m$ of observed values will become

$$y \approx Aw + be,$$

where w and b be the solution of (5).

By introducing the Lagrange multipliers $u_1 = (u_{11},\ldots,u_{1m})^t$ and $u_2 = (u_{21},\ldots,u_{2m})^t$ in R^m the Lagrangian function L can be obtained to be:

$$\begin{aligned} L(w,b,\xi,\xi^*,u_1,u_2) = {} & \frac{1}{2}(w^t w + b^2) + \frac{\nu}{2}\sum_{i=1}^m (\xi_i^2 + \xi_i^{*2}) \\ & + \sum_{i=1}^m u_{1i}(y_i - A_i w - b - \varepsilon - \xi_i) \\ & + \sum_{i=1}^m u_{2i}(A_i w + b - y_i - \varepsilon - \xi_i^*). \end{aligned}$$

Using the condition that the partial derivatives of L with respect to the primal variables will be zero at optimality the dual problem can be written as a minimization problem of the following form (Musicant & Feinberg, 2004)

$$\begin{aligned} \min_{0\leqslant u_1,u_2\in R^m} & \frac{1}{2}\left[(u_1-u_2)^t(AA^t + ee^t)(u_1-u_2)\right] + \frac{1}{2\nu}(u_1^t u_1 + u_2^t u_2) \\ & - y^t(u_1-u_2) + \varepsilon e^t(u_1+u_2), \end{aligned} \tag{6}$$

in which

$$w = A^t(u_1-u_2) \text{ and } b = e^t(u_1-u_2) \tag{7}$$

hold. Now from (1) and (7) the linear regression estimation function $f(\cdot)$ is given by