



Lagrangian support vector regression via unconstrained convex minimization



S. Balasundaram*, Deepak Gupta, Kapil

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

ARTICLE INFO

Article history:

Received 31 March 2013
Received in revised form 4 September 2013
Accepted 4 December 2013

Keywords:

Generalized derivative approach
Smooth approximation
Support vector regression
Unconstrained convex minimization

ABSTRACT

In this paper, a simple reformulation of the Lagrangian dual of the 2-norm support vector regression (SVR) is proposed as an unconstrained minimization problem. This formulation has the advantage that its objective function is strongly convex and further having only m variables, where m is the number of input data points. The proposed unconstrained Lagrangian SVR (ULSVR) is solvable by computing the zeros of its gradient. However, since its objective function contains the non-smooth 'plus' function, two approaches are followed to solve the proposed optimization problem: (i) by introducing a smooth approximation, generate a slightly modified unconstrained minimization problem and solve it; (ii) solve the problem directly by applying generalized derivative. Computational results obtained on a number of synthetic and real-world benchmark datasets showing similar generalization performance with much faster SVR learning speed in accordance with the conventional SVR and training time very close to least squares SVR clearly indicate the superiority of ULSVR solved by smooth and generalized derivative approaches.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Support Vector Machines (SVMs) are state-of-the-art machine learning techniques based on Vapnik and Chervonenkis structural risk minimization principle (Vapnik, 2000). They are computationally powerful tools showing better generalization ability compared to other machine learning methods on a wide variety of real-world problems like face detection (Osuna, Freund, & Girosi, 1997), gene prediction (Guyon, Weston, Barnhill, & Vapnik, 2002), text categorization (Joachims, Ndellec, & Rouveriol, 1998) and image segmentation (Chen & Wang, 2005). It is well known that SVM formulation (Cristianini & Shawe-Taylor, 2000; Vapnik, 2000) can be posed as a quadratic programming problem (QPP) with linear inequality constraints which is a convex programming problem having a unique optimal solution. Although SVM has been initially proposed for classification, later it has been extended to regression (Vapnik, 2000).

The function approximation by regression is of great importance in many fields of research such as bioinformatics, control theory, economics, information science and signal processing. The main challenge in developing a useful regression model is to capture accurately the underlying functional relationship between the

given inputs and their output values. Once the resulting model is obtained it can be used as a tool for analysis, simulation and prediction.

For a given set of data points, support vector regression (SVR) aims at determining a linear regressor where the input data are either considered in the original pattern space itself or taken into a higher dimensional feature space. With the introduction of ϵ -insensitive error loss function proposed by Vapnik (2000), SVR has emerged as a powerful paradigm of choice for regression because of its improved generalization performance in comparison to other popular machine learning methods like artificial neural networks.

Considering the minimization of the square of 2-norm of the slack variables 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 parallel planes, Mangasarian and Musicant (2001a, 2001b) studied "equivalent" SVM formulations for classification problems resulting in positive-definite dual problems having only non-negative constraints. For the formulation of SVM as an unconstrained optimization problem whose objective function becomes not twice differentiable, a smoothing technique is adopted to derive a new SVM formulation called Smooth SVM (SSVM) in Lee and Mangasarian (2001). As its extension to ϵ -insensitive error loss based SVR, termed Smooth SVR (SSVR), we refer the reader to Lee, Hsieh, and Huang (2005). Also, on a finite Newton method of solution for SVM for classification with generalized Hessian approach, see Fung and Mangasarian (2003) and further on its extension to SVR, the interested reader is referred to Balasundaram and Kapil (2011). In addition, for the extension of

* Corresponding author. Tel.: +91 11 26704724; fax: +91 11 26741586.

E-mail addresses: balajnu@gmail.com, bala_jnu@hotmail.com (S. Balasundaram), deepakjnu85@gmail.com (D. Gupta), navkapil@gmail.com (Kapil).

the Active set SVM (ASVM) proposed for the classification (Mangasarian & Musicant, 2001b) to SVR, we refer Musicant and Feinberg (2004). For a review on optimization techniques used for training SVMs, see Shawe-Taylor and Sun (2011). Finally on the interesting work on multitask learning and multiview learning methods as extension of kernel based methods, we refer Ji and Sun (2013), Sun (2011) and Xie and Sun (2012).

It is well-known that the conventional ε -insensitive SVR is formulated as a convex quadratic minimization problem having $2m$ nonnegative variables and $2m$ linear inequality constraints whereas 2-norm SVR introduces only $2m$ linear inequality constraints, where m is the number of training points. In either case, more variables and constraints enlarge the problem size and therefore increase in the computational cost for solving the regression problem. Recently, a new nonparallel plane regressor termed as twin SVR (TSVR) is proposed in Peng (2010a) wherein a pair of QPPs, each having m number of constraints, is solved instead of solving a single QPP having $2m$ constraints as in the conventional SVR which makes TSVR works faster than SVR.

In our approach, the 2-norm SVR in dual is reformulated as a single, unconstrained minimization problem having only m variables. We term such reformulation as unconstrained Lagrangian SVR (ULSVR). The proposed ULSVR model can be solved using gradient based methods. However, since the objective function contains a term having non-smooth 'plus' function, two approaches are taken to solve the minimization problem: (i) two smooth approximation functions, studied in Lee and Mangasarian (2001) and Peng (2010b), are introduced and their slightly modified unconstrained minimization problems are solved; (ii) directly solve the minimization problem by applying generalized derivative. In both the approaches, by applying simple iterative methods, the zeros of the gradient are computed.

Finally, for the study of Lagrangian and implicit Lagrangian SVR solved using block matrices with the advantage of reduced training cost, see the work of Balasundaram and Kapil (2010, 2011).

The effectiveness of the proposed approach in formulating ULSVR problem and solving it using various iterative methods is demonstrated by performing numerical experiments on a number of synthetic and real-world benchmark datasets and comparing their results with conventional SVR and least squares SVR. Comparable generalization performance in solving ULSVR with less computational training time clearly indicates its suitability on problems of interest.

The paper is organized as follows. In Section 2, formulations of the conventional and 2-norm SVR are introduced. Section 3 briefly dwells on least squares SVR. The proposed ULSVR formulation having only m variables and various algorithms used for solving it are described in Section 4. On a number of synthetic and real-world datasets numerical experiments are performed and the results obtained by ULSVR are compared with SVR and least squares SVR in Section 5. Finally we conclude our work in Section 6.

In this work, all vectors are considered as column vectors. For any two vectors x, y in \mathfrak{R}^n , their inner product will be denoted by $x^t y$ where x^t is the transpose of x . When x is orthogonal to y we write $x \perp y$. The 2-norm of a vector x will be denoted by, $\|x\|$. For $x = (x_1, \dots, x_n)^t \in \mathfrak{R}^n$, the plus function x_+ is defined as: $(x_+)_i = \max\{0, x_i\}$, where $i = 1, \dots, n$. Further we define the step function x_* as: $(x_*)_i = 1$ for $x_i > 0$, $(x_*)_i = 0$ if $x_i < 0$ and $(x_*)_i = 0.5$ when $x_i = 0$. The identity matrix of appropriate size is denoted by I and the diagonal matrix of order n whose diagonal elements become the components of the vector $x \in \mathfrak{R}^n$ by $\text{diag}(x)$. The column vector of ones of dimension m is denoted by e . If f is a real valued function of the variable $x = (x_1, \dots, x_n)^t \in \mathfrak{R}^n$ then its gradient vector and Hessian matrix are defined by, $\nabla f = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})^t$ and $\nabla^2 f = (\partial^2 f / \partial x_i \partial x_j)_{i,j=1,\dots,n}$ respectively.

2. Support vector regression (SVR)

In this section, we briefly describe the conventional SVR and 2-norm SVR formulations.

Assume that a set of input samples $\{(x_i, y_i)\}_{i=1,2,\dots,m}$ be given where for each training example $x_i \in \mathfrak{R}^n$, let $y_i \in \mathfrak{R}$ be its corresponding observed value. Further, let the training examples be represented by a matrix $A \in \mathfrak{R}^{m \times n}$ whose i th row is defined to be the row vector x_i^t and the vector of observed values be denoted by $y = (y_1, \dots, y_m)^t$.

The primary goal of SVR is in seeking a nonlinear regression function by mapping the training examples into a higher dimensional feature space via a nonlinear mapping $\varphi(\cdot)$ and learning a linear regressor in the feature space. Assume that the nonlinear regression estimating function $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ takes the form:

$$f(x) = w^t \varphi(x) + b,$$

where w is a vector in the feature space and b is a scalar threshold.

The ε -insensitive SVR formulation proposed by Vapnik (2000) aims at determining w and b as the solution of the following QPP (Cristianini & Shawe-Taylor, 2000; Vapnik, 2000)

$$\min_{w, b, \xi_1, \xi_2} \frac{1}{2} w^t w + C(e^t \xi_1 + e^t \xi_2)$$

subject to

$$y_i - w^t \varphi(x_i) - b \leq \varepsilon + \xi_{1i}, \quad w^t \varphi(x_i) + b - y_i \leq \varepsilon + \xi_{2i} \\ \text{and } \xi_{1i}, \xi_{2i} \geq 0 \quad \text{for } i = 1, 2, \dots, m, \quad (1)$$

where $\xi_1 = (\xi_{11}, \dots, \xi_{1m})^t$, $\xi_2 = (\xi_{21}, \dots, \xi_{2m})^t$ are vectors of slack variables and $C > 0$, $\varepsilon > 0$ are input parameters. Other than the free variables w and b , this formulation has $2m$ nonnegative variables and $2m$ linear inequality constraints.

In practice, rather than solving the primal problem (1), its dual is solved. In fact, replacing the dot product $\varphi(x_i)^t \varphi(x_j)$ by a kernel function $k(x_i, x_j)$, the dual of problem (1) can be written as:

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

subject to

$$\sum_{i=1}^m (u_{1i} - u_{2i}) = 0 \quad \text{and} \quad 0 \leq u_1, u_2 \leq Ce, \quad (2)$$

where $u_1 = (u_{11}, \dots, u_{1m})^t$ and $u_2 = (u_{21}, \dots, u_{2m})^t$ in \mathfrak{R}^m are Lagrange multipliers.

Assuming that the kernel mapping is given, using the solution of the dual problem (2), the nonlinear decision function $f(\cdot)$ is obtained (Cristianini & Shawe-Taylor, 2000; Vapnik, 2000) as: for any $x \in \mathfrak{R}^n$,

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

Though different kernel functions such as polynomial or Gaussian radial basis functions can be used, the popular kernel being the Gaussian kernel and is defined as: for vectors $x, z \in \mathfrak{R}^n$,

$$k(x, z) = \exp(-\mu \|x - z\|^2), \quad (3)$$

where $\mu > 0$ is a parameter.

Following the approach of Mangasarian and Musicant (2001a, 2001b) and Musicant and Feinberg (2004), by considering the square of the 2-norm of the slack variables ξ_1, ξ_2 instead of 1-norm

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات