Knowledge-Based Systems 30 (2012) 87-94

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Multivariate convex support vector regression with semidefinite programming

Yongqiao Wang*, He Ni

School of Finance, Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China

ARTICLE INFO

Article history: Received 25 August 2011 Received in revised form 21 November 2011 Accepted 19 December 2011 Available online 27 December 2011

Keywords: Support vector regression Shape-restriction Convexity Semidefinite programming Linear matrix inequality constraints

1. Introduction

Avoiding strong prior assumptions on functional forms, nonparametric regression methods are powerful tools for data description and exploration. The major advantage of these methods is that they do not require analysts to explicitly specify a given parametric structure on the data; instead, the data are allowed to "speak for themselves". Nonparametric regression methods are attracting increasing attention from many areas [45,10].

Even though an analyst may not know the exact form of the relationship, he/she usually has some prior knowledge of its shape, such as monotone and convex. Typical examples appear in economics (utility function, production or cost functions), medicine (dose response experiments) or biology (growth curves). For a rational consumer, his/her utility function is widely recognized to be non-decreasing and concave. By fitting an explicitly specified function, parametric methods can certainly obtain estimation that satisfies prior-known shape restrictions. But parametric methods are vulnerable to model specification error. It is widely recognized by nonparametric statisticians that shape-restricted nonparametric regression can better predict the relationship between predictors and responses.

Shape-restricted nonparametric regression dates back to seminal works [15,7]. The first paper dealt with least squares estimation of a concave function, while the second discussed the estimation of monotone functions. Since then, a lot of results on different

ABSTRACT

As one of important nonparametric regression method, support vector regression can achieve nonlinear capability by kernel trick. This paper discusses multivariate support vector regression when its regression function is restricted to be convex. This paper approximates this convex shape restriction with a series of linear matrix inequality constraints and transforms its training to a semidefinite programming problem, which is computationally tractable. Extensions to multivariate concave case, ℓ_2 -norm Regularization, ℓ_1 and ℓ_2 -norm loss functions, are also studied in this paper. Experimental results on both toy data sets and a real data set clearly show that, by exploiting this prior shape knowledge, this method can achieve better performance than the classical support vector regression.

© 2011 Elsevier B.V. All rights reserved.

shape-restricted nonparametric regression methods have been published. One can refer to [8] and, more recently, [49] for the literature on isotonic regression, i.e. monotone regression. In case of convex or concave regression, some statistical properties of concave regression [15], such as consistency, rate of convergence and asymptotic distribution, have been analyzed by [13,23,12]. There is vast literature on shape-restricted nonparametric regression. The mainly concerned shapes include monotonicity, convexity, concavity, super-modularity, unimodality, etc. Alternative to least-squares minimization, many spline, kernel smoothing and wavelets-based techniques have been applied in shape-restricted nonparametric regression.

This paper focuses on estimating a multivariate regression function when it is known to be convex or concave. The extension from univariate convex nonparametric regression to multivariate convex regression is not straightforward. For univariate nonparametric regression, the convex shape restriction requires its second-order differential to be non-negative. While in *d*-dimension $(d \ge 2)$ multivariate nonparametric regression, this convex shape restriction is equivalent to positive semidefinite requirements of Hessian matrix on its domain. Though univariate convex nonparametric regression has been extensively researched, there are few work on multivariate case, except [2,26,27,1,17,33]. Matzkin [26,27] considered the case where the variable of interest was discrete. In [2,1], the regression function was assumed to be contaminated by errors with specified distributions, so it could be obtained by maximum likelihood estimation. Instead of specifying error distributions, [17,33] acquired regression function by least squares minimization. The most important obstacle of the above methods is that they can only obtain piece-linear surface, which is not differentiable at knots.





^{*} Corresponding author. Tel.: +86 571 28877720; fax: +86 571 28877705. *E-mail addresses*: wangyq@zjgsu.edu.cn (Y. Wang), nihe@zjgsu.edu.cn (H. Ni).

^{0950-7051/\$ -} see front matter © 2011 Elsevier B.V. All rights reserved. doi:10.1016/j.knosys.2011.12.010

This paper proposes an alternative nonparametric estimation of multivariate convex or concave regression function, which is based on support vector regression [35]. Support vector regression belongs to kernels methods [32,34], which have found successful applications in many areas such as credit risk evaluation [44,21], fraud detection [28] and time series forecasting[47,46]. To achieve nonlinear regression, support vector regression forms an optimal linear regressive hyperplane by a mapping from the input space to a higher-dimension feature space. Support vector regression has great advantage over [2,1,17,33], because its regression function is smooth and differentiable everywhere.

Incorporating qualitative prior shape knowledge into support vector regression has been explored by [29,39]. [29] built monotone least squares support vector machine, which imposed the monotonicity-related constraints on every pair monotone samples. [39] also analyzed support vector regression when the derivatives of the regression function was restricted to be boundary, which included monotone, convex and concave shape restrictions. But [39] can not be generalized to bivariate or multivariate convex and concave cases, which needs semidefinite constraints, instead of common linear inequality constraints. It is also widely known in machine learning and neural networks area that one can achieve better performance by incorporating prior knowledge [19]. Literature on prior knowledge based support vector machine includes [11,29,25,20].

This paper employs semidefinite programming to solve convex or concave multivariate support vector regression. Applications of semidefinite programming also can be found in [18]. In our method, the convex shape restriction, approximated by a series of linear matrix inequality constraints on every training points, can force the regression function to be convex or concave. By this way, this method uses this prior shape knowledge on the function between predictors and responses. We expect that this exploitation of prior qualitative shape knowledge can improve out-of-sample regression performance. It is obvious that this method has two novelties. First, compared with parametric methods, its nonparametric characteristics can efficiently avoid model specification error. Second, maximization of the additional regularization term enables the method to minimize not only the empirical error, but also the generalization error.

The paper is organized as follows. Section 2 introduces the main idea of solving multivariate convex or concave support vector regression by semidefinite programming. The capability of our method is verified by two artificial data sets and one real data set in Section 3. Other variants, including loss functions and regularization terms, are analyzed in Section 4. The paper is concluded in Section 5.

All vectors are column vectors written in boldface and lowercase letters whereas matrices are boldface and uppercase, except for the *i*th row of a matrix **A** that is denoted **A**_{*i*}. The vectors **0** and **1** are vectors of appropriate dimensions with all their components respectively equal to 0 and 1. **I** is the identity matrix with appropriate size. The matrix $\mathbf{X} \in \mathbb{R}^{N \times d}$ contains all the training samples \mathbf{x}_i , i = 1, ..., N, as rows. The vector $\mathbf{y} \in \mathbb{R}^N$ contains all the target values y_i for these samples. $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the kernel function. For $\mathbf{A} \in \mathbb{R}^{m \times d}$ and $\mathbf{B} \in \mathbb{R}^{n \times d}$ containing *d*-dimensional sample vectors, the kernel **K** (**A**, **B**) maps $\mathbb{R}^{m \times d} \times \mathbb{R}^{n \times d}$ to $\mathbb{R}^{m \times n}$ with **K**(**A**, **B**)_{*i*, *j*} = *k*(*A*_{*i*}, *B*_{*j*}). The kernel matrix **K**(**X**, **X**) will be written **K** for short. For a vector **v**, diag(**v**) is the diagonal matrix with the components of **v** on its diagonal. For a collection of symmetric matrices **A**₁,..., **A**_{*i*}, diag(**A**₁,..., **A**_{*i*})

$$\operatorname{diag}(\mathbf{A}_{1},\ldots,\mathbf{A}_{i}) = \begin{pmatrix} \mathbf{A}_{1} & & \\ & \ddots & \\ & & \mathbf{A}_{i} \end{pmatrix}.$$
 (1)

、

. -

For any symmetric square matrix \mathbf{A} , $\mathbf{A} \succeq 0$ means that \mathbf{A} is positive semidefinite and $\mathbf{A} \preceq 0$ means that \mathbf{A} is negative semidefinite.

2. Multivariate convex support vector regression

2.1. Support vector regression

Let samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^N \subset \mathcal{X}_i \times \mathbb{R}$ be the training data, where $\mathcal{X} \subset \mathbb{R}^d$ is a closed and convex set. The object of nonparametric regression is to find a function $f : \mathcal{X} \to \mathbb{R}, \mathbf{x} \to f(\mathbf{x})$ that minimizes the objective

$$\min_{f \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} L(f(\mathbf{x}_i), y_i) + C \|f\|_{K}^{2},$$
(2)

where $L(\cdot, \cdot)$ denotes the chosen loss function, $||f||_K$ is the norm in the reproducing kernel Hilbert space (\mathcal{H}) defined by the kernel $k(\mathbf{x}, \mathbf{x}')$, C is a free parameter introduced to tune the trade-off between the empirical error minimization and the regularization term maximization. A large C might make the model not sufficiently flexible to explain the data, while a small C might cause overfitting. According to [43], the framework (2) includes the entire family of smoothing splines, additive and interaction spline models.

To make the above estimation implementable, two components should be explicitly specified. The first is the loss function *L*. Typical choices of *L* include:

*l*₁-norm loss. This loss puts the most weight on small residuals and the least weight on large residuals. This penalty allows less either zero or very small residuals but relatively more large residuals.

$$\mathcal{L}(f(\mathbf{x}_i), y_i) = |f(\mathbf{x}_i) - y_i|.$$
(3)

*t*₂-norm loss. This loss puts very small weight on small residuals, but strong weight on large residuals. This penalty results more modest residuals but relatively less larger ones.

$$L(f(\mathbf{x}_i), \mathbf{y}_i) = (f(\mathbf{x}_i) - \mathbf{y}_i)^2.$$
(4)

• ϵ -insensitive loss. This loss is also known as deadzone-linear penalty. This loss puts no weight on small residuals. For this one, many residuals have value $\pm \epsilon$, i.e., right at the edge of the free zone, for which no penalty is applied. When $\epsilon = 0$, it is equivalent to ℓ_1 -norm loss.

$$L(f(\mathbf{x}_i), y_i) = |f(\mathbf{x}_i) - y_i|_{\epsilon},$$
(5)

where $|\xi|_{\epsilon}$ is defined as

$$|\xi|_{\epsilon} = \begin{cases} 0 & \text{if } |\xi| \leq \epsilon, \\ |\xi| - \epsilon & \text{otherwise.} \end{cases}$$
(6)

It is widely known [37,42] that, tolerating a small error in fitting, i.e. disregarding errors that fall within some ϵ , can improve performance over zero-tolerance loss function, such as ℓ_1 and ℓ_2 norm loss. This section only utilizes ϵ -insensitive loss. Extensions to other loss functions will be explored in Section 4.

Another component should be specified is the kernel function *k*. Two popular choices of $k(\cdot, \cdot)$ in practice are:

• Polynomial kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \mathbf{x}_i' \mathbf{x}_j\right)^p, \quad p \in \mathbb{N}.$$
(7)

• Gaussian or Radial Basis Function (RBF) kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right\}, \quad \sigma^2 > 0.$$
(8)

دريافت فورى 🛶 متن كامل مقاله

- 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
- ISIArticles مرجع مقالات تخصصی ایران