



Learning errors of linear programming support vector regression [☆]

Feilong Cao ^{*}, Yubo Yuan

Institute of Metrology and Computational Science, China Jiliang University, Hangzhou 310018, Zhejiang Province, China

ARTICLE INFO

Article history:

Received 1 June 2009
 Received in revised form 18 September 2010
 Accepted 4 October 2010
 Available online 13 October 2010

Keywords:

Regression
 Support vector machine
 Linear programming
 Generalization error
 Empirical error

ABSTRACT

In this paper, we give several results of learning errors for linear programming support vector regression. The corresponding theorems are proved in the reproducing kernel Hilbert space. With the covering number, the approximation property and the capacity of the reproducing kernel Hilbert space are measured. The obtained result ([Theorem 2.1](#)) shows that the learning error can be controlled by the sample error and regularization error. The mentioned sample error is summarized by the errors of learning regression function and regularizing function in the reproducing kernel Hilbert space. After estimating the generalization error of learning regression function ([Theorem 2.2](#)), the upper bound ([Theorem 2.3](#)) of the regularized learning algorithm associated with linear programming support vector regression is estimated.

Crown Copyright © 2010 Published by Elsevier Inc. All rights reserved.

1. Introduction

The main aim of this paper is the error analysis for the linear programming support vector regression (SVR) problem in learning theory. To this end, this paper is organized as follows. We start by giving a brief introduction of the basic techniques of support vector machine (SVM) or SVR in Section 1, including the history, motivation, most important publications and some corresponding results with this paper. Moreover, a brief review of linear programming support vector regression (LP-SVR) and quadratic programming one (QP-SVR) in the reproducing kernel Hilbert space is presented in Section 2. Learning error analysis of LP-SVR is given in Section 3. In Section 4, we plus a short conclusion and summary with our research work, also, some future works are attached.

First of all, we briefly describe the historic background of support vector learning algorithm (SVLA). The SVLA is a non-linear generalization of the generalized portrait algorithm developed in the sixties (see, [1,2]).

As such, the SVLA is firmly grounded in the framework of statistical learning theory. In 1995, Cortes and Vapnik (see, [3]) proposed a learning network named support-vector network. Originally, it is a learning machine for two-group classification problems. One of the most important idea is that input vectors had been non-linearly mapped to a very high-dimension feature space. In this feature space a linear decision surface was constructed. They showed that the training data (separable and non-separable) could be separated without errors. After that, this learning network became more and more popular and renamed as SVM.

Today, SVM has become an important subject in learning theory and has evolved into an active area of research. Mathematically, SVM is from a pattern classification problem based on a given classification of m points $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ in

[☆] The research was supported by the National Natural Science Foundation of China (Nos. 60873206, 61001200) and the Natural Science Foundation of Zhejiang Province of China (No. Y7080235).

^{*} Corresponding author. Tel.: +86 571 86835737.

E-mail address: feilongcao@gmail.com (F. Cao).

the n -dimensional space \mathbb{R}^n , represented by an $m \times n$ matrix $\mathbf{A} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)^T$, given the membership of each data point \mathbf{x}_i , $i = 1, 2, \dots, m$ in the classes 1 or -1 as specified by a given $m \times m$ diagonal matrix D with 1 or -1 diagonals.

Primarily, SVM (see, [4]) is given by the following quadratical programming with linear inequalities constraints

$$\begin{aligned} \min_{(\alpha, b) \in \mathbb{R}^{n+1}} & \quad \frac{1}{2} \|\alpha\|^2 + \frac{1}{2} \|\mathbf{y}\|^2, \\ \text{s.t.} & \quad D(\mathbf{A}\alpha + b\mathbf{e}) \geq \mathbf{e} + \mathbf{y}, \\ & \quad \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{1}$$

Model (1) can be seen as the original model of QP-SVM. Here, α is a vector of separator coefficients (direction vector of classification hyperplane), b is an offset (the control parameter of the distance of hyperplane plane to the origin) and $\mathbf{e} \in \mathbb{R}^m$ stands for a vector of ones.

The decision function of classification is given by

$$f(\mathbf{x}) = \text{sign}(\alpha^T \mathbf{x} + b). \tag{2}$$

By now, many different forms of QP-SVM (1) have been introduced for different purposes (see [4]). In this work, we mainly pay our attention to the learning error or the convergence rate of the proposed algorithm. For the convergence rate of QP-SVM (1), there are many works. We refer the readers to Steinart [5], Zhang [6], Wu and Zhou [7], Zhao and Yin [8], Hong [9], and Wu et al. [10].

Among of forms of SVM, the LP-SVM is important because of its linearity and flexibility for large data setting. Many authors have introduced the LP-SVM. We refer the readers to Bradley and Mangasarian [11], Kecman and Hadaic [12], Niyogi and Girosi [13], Pedroso and Murata [14] and Vapnik [4]. Its primal optimization model is as follows

$$\begin{aligned} \min_{(\lambda, \mathbf{y}) \in \mathbb{R}^{2m}} & \quad \frac{1}{m} \mathbf{e}^T \lambda + \frac{1}{C} \mathbf{e}^T \mathbf{y}, \\ \text{s.t.} & \quad D(\mathbf{A}\mathbf{A}^T \lambda + b\mathbf{e}) \geq \mathbf{e} - \mathbf{y}, \\ & \quad \lambda \geq \mathbf{0}, \quad \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{3}$$

The trade-off factor $C = C(m) > 0$ depends on m and is crucial.

Many experiments demonstrate that LP-SVM is efficient and perform even better than QP-SVM for some purposes: capable of solving huge sample size problems (see, [11]), improving the computational speed (see, [14]), and reducing the number of support vectors (see, [12]).

However, little is known for the learning error or the convergence of the LP-SVM. We only find that a classification problem of LP-SVM was studied in Wu and Zhou [15].

The primal goal of this paper is to address in the investigation of regression problem and to provide the error analysis for linear programming support vector regression (LP-SVR).

2. LP-SVR and QP-SVR in the reproducing kernel Hilbert space

For regression problem, SVR can be seen as a SVLA with continuous output or support vector machines for function estimation. In 2004, Smola and Schölkopf [16] have given an overview of the basic ideas underlying SVM for function estimation. They indicated that the SVR was more difficult to analysis the learning error and convergence. This is the motivation of this work.

Let (X, d) be a compact metric space and let $X \subset \mathbb{R}^n$, $Y = \mathbb{R}$. In this work, we only discuss the single output regression problem. For multi-output regression problem, there is some difficult coming from the multivariate output space. It seems that it is a very challenge work to study learning errors for multi-output LP-SVR in reproducing Hilbert spaces. In 2009, Liu et al. [17] studied the output space as a Riemannian submanifold to incorporate its geometric structure into the regression process. They proposed a locally linear transformation (LLT), to define the loss functions on the output manifold. An algorithm was given under the framework of SVR.

Let ρ be a probability distribution on $S = X \times Y$. The error (or generalization error) for a function $f: X \rightarrow Y$ is defined as

$$\mathcal{E}(f) = \int_S (y - f(x))^2 d\rho.$$

The function that minimizes the error is called the regression function. It is given by

$$f_\rho(x) = \int_Y y d\rho(y|x), \quad \forall x \in X,$$

where $\rho(\cdot|x)$ is the conditional probability measure at x induced by ρ .

The target of regression problem is to learn the regression function or to find good approximation from random samples.

The least-square algorithm for the regression problem is a discrete least-square problem associated with a Mercer kernel. Let $K: X \times X \rightarrow \mathbb{R}$ be continuous, symmetric and positive semidefinite, i.e., for any finite set of distinct points

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