Neurocomputing 140 (2014) 41–52

Contents lists available at ScienceDirect

Neurocomputing

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

A robust least squares support vector machine for regression and classification with noise

Xiaowei Yang [a,*], Liangjun Tan [a], Lifang He [b]

[a] Department of Mathematics, School of Sciences, South China University of Technology, Guangzhou 510641, PR China
[b] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, PR China

ARTICLE INFO

Article history:
Received 18 September 2013
Received in revised form
12 March 2014
Accepted 15 March 2014
Communicated by X. Gao
Available online 13 April 2014

Keywords:
Least squares support vector machines
Weighted least squares support vector machines
Robust least squares support vector machine
Regression
Classification
Noise

ABSTRACT

Least squares support vector machines (LS-SVMs) are sensitive to outliers or noise in the training dataset. Weighted least squares support vector machines (WLS-SVMs) can partly overcome this shortcoming by assigning different weights to different training samples. However, it is a difficult task for WLS-SVMs to set the weights of the training samples, which greatly influences the robustness of WLS-SVMs. In order to avoid setting weights, in this paper, a novel robust LS-SVM (RLS-SVM) is presented based on the truncated least squares loss function for regression and classification with noise. Based on its equivalent model, we theoretically analyze the reason why the robustness of RLS-SVM is higher than that of LS-SVMs and WLS-SVMs. In order to solve the proposed RLS-SVM, we propose an iterative algorithm based on the concave–convex procedure (CCCP) and the Newton algorithm. The statistical tests of the experimental results conducted on fourteen benchmark regression datasets and ten benchmark classification datasets show that compared with LS-SVMs, WLS-SVMs and iteratively reweighted LS-SVM (IRLS-SVM), the proposed RLS-SVM significantly reduces the effect of the noise in the training dataset and provides superior robustness.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Support vector machines (SVMs) are very important methodologies for classification [1–4] and regression [5–7] in the fields of pattern recognition and machine learning. It has been widely applied to many real world pattern recognition problems, such as text classification [8,9], image classification [10,11], feature extraction [12–14], web mining [15] and function estimation [16,17]. Based on equality constraints instead of inequality ones, two least squares support vector machines (LS-SVMs) are proposed for classification [18,19] and regression [20,21], respectively. Recently, a matrix pattern based LS-SVM is also presented [22]. The solutions of LS-SVMs are obtained by solving a set of linear equations instead of solving a quadratic programming (QP) problem as in SVM. Several effective numerical algorithms have been suggested, such as the conjugate gradient based iterative algorithm [19,21,23,24], the reduced set of linear equations based algorithm [25], the sequential minimal optimization algorithm (SMO) [26], and the Sherman–Morrison–Woodbury (SMW) identity based algorithm [27].

At present, LS-SVMs have been widely applied to text classification [28], image processing [29–31], time series forecasting [21,32,33], and control [34–36]. Unfortunately, in real-world applications, there exist two main drawbacks in LS-SVMs. The first one is their solutions are non-sparse [37,38] and the second one is their training processes are sensitive to noise in the training dataset due to over-fitting [39,40]. In order to deal with the first problem, some pruning algorithms have been proposed [41–44]. In order to deal with the second problem, two weighted LS-SVMs (WLS-SVMs) have been presented for regression [45] and classification [46], respectively. A key issue for WLS-SVMs is how to assign suitable weights to training samples. In the previous studies, the weights are assigned to the training samples by a two-stage method [45,47] and a multi-stage method [48]. Theoretical analyses and the related experiments show that WLS-SVMs are robust to some noise.

In the field of machine learning, robust loss function is usually one of the key issues in designing a robust algorithm. At present, various margin-based loss functions, such as squared loss, logistic loss, hinge loss, exponential loss, 0-1 loss, and brownboost loss, have been used to search for the optimal classification and regression functions. From their function curves [49], we know that squared loss, logistic loss, hinge loss, and exponential loss are the upper boundary on the generalization error of a 0-1 loss. When the training sample has a large negative margin, squared loss, hinge loss and exponential loss are larger than brownboost loss. Therefore, the brownboost loss is usually more robust than the other loss functions. Recently, motivated by the link between

* Corresponding author. Tel.: +86 20 87110446.
E-mail address: xwyang@scut.edu.cn (X. Yang).

http://dx.doi.org/10.1016/j.neucom.2014.03.037
0925-2312/© 2014 Elsevier B.V. All rights reserved.

the pinball loss and quantile regression, Huang et al. introduced the pinball loss to classification problems and proposed the pinball loss SVM (pin-SVM) [50]. The theoretical analysis and the experimental results show that compared to the hinge loss SVM, the pin-SVM is less sensitive to the feature noise around the decision boundary and more stable for re-sampling.

In order to avoid setting the weights of the training samples, which greatly influence the robustness of WLS-SVMs, in this study, inspired by the ideas in [51], we propose a novel robust LS-SVM (RLS-SVM) based on the truncated least squares loss function for regression and classification with noise. Based on the definition of influence function [52], we show that the proposed loss function is insensitive to noise. Considering that the proposed loss function is neither differentiable nor convex, inspired by [53], we firstly give a smoothing procedure to make the proposed loss function smooth. Secondly, using the concave–convex procedure (CCCP) [54], we transform solving a concave–convex optimization problem into solving iteratively a series of the convex optimization problems. Finally, we apply the Newton algorithm [53] to solve these convex optimization problems. In order to test the robustness of RLS-SVM, we conduct a set of experiments on four synthetic regression datasets, fourteen benchmark regression datasets, two synthetic classification datasets and ten benchmark classification datasets. In the analysis of the experimental results, the Wilcoxon signed-ranks test and the Friedman test [55] are used to check the significant of RLS-SVM.

This paper is organized as follows. In Section 2, we briefly review LS-SVMs and WLS-SVMs. In Section 3, we propose RLS-SVM. In Section 4, we theoretically analyze the reason why the robustness of RLS-SVM is higher than that of LS-SVMs and WLS-SVMs. An algorithm for RLS-SVM is given based on the CCCP and the Newton algorithm in Section 5. The experimental results and analyses are presented in Section 6. Finally, conclusions are given in Section 7.

2. Least squares support vector machines and weighted least squares support vector machines

2.1. Least squares support vector machine for regression

Considering a training set of l pairs of samples $\{\mathbf{x}_i, y_i\}_{i=1}^{l}$ for regression problem, where $\mathbf{x}_i \in R^n$ are the input data and $y_i \in R$ are the corresponding prediction values, LS-SVM for the regression problem is a QP problem based on the equality constraints and can be described in the following [20,21]:

$$\min_{\mathbf{w},b,\xi} J(\mathbf{w}, b, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2, \tag{1}$$

s. t.

$$y_i - [\mathbf{w}^T\varphi(\mathbf{x}_i) + b] = \xi_i, \quad i = 1, \cdots, l, \tag{2}$$

where $\mathbf{w}$ is the normal of the hyperplane, ξ_i is the error of the ith training sample, $\varphi(\mathbf{x}_i)$ is a nonlinear function that maps $\mathbf{x}_i$ to a high-dimensional feature space, C is a regularized parameter balancing the tradeoff between the margin and the error, and b is a bias.

The Lagrangian function of the optimization problem (1) and (2) is

$$L(\mathbf{w}, b, \xi, \boldsymbol{\alpha}) = J(\mathbf{w}, b, \xi) - \sum_{i=1}^{l}\alpha_i\{\mathbf{w}^T\varphi(\mathbf{x}_i) + b + \xi_i - y_i\}, \tag{3}$$

where α_i are the Lagrangian multipliers.

The optimal conditions can be written as the following system of linear equations:

$$\begin{pmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \boldsymbol{\Omega} + \frac{\mathbf{I}}{C} \end{pmatrix}\begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{Y} \end{pmatrix}, \tag{4}$$

where $\mathbf{I} \in R^{l \times l}$ is an identity matrix,

$$\mathbf{Y} = (y_1, y_2, \ldots, y_l)^T, \tag{5}$$

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)^T, \tag{6}$$

$$\mathbf{e} = (1, 1, \ldots, 1)^T, \tag{7}$$

$$\boldsymbol{\Omega} = (\Omega_{ij}) = (k(\mathbf{x}_i, \mathbf{x}_j)), \tag{8}$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle\varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j)\rangle. \tag{9}$$

2.2. Weighted least squares support vector machine for regression

WLS-SVM for the regression problem is described in the following [45]:

$$\min_{\mathbf{w},b,\xi} J(\mathbf{w}, b, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{l}s_i\xi_i^2, \tag{10}$$

s. t.

$$y_i - (\mathbf{w}^T\varphi(\mathbf{x}_i) + b) = \xi_i, \quad i = 1, \cdots, l, \tag{11}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_l)$ is a vector of weights associated with the training samples. If $s_j = 0$, one can delete the corresponding training sample from the model. The optimal dual variables can be given by the solution of the following system of linear equations:

$$\begin{pmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \boldsymbol{\Omega} + \frac{1}{C}diag\left(\frac{1}{s_1}, \frac{1}{s_2}, \ldots, \frac{1}{s_l}\right) \end{pmatrix}\begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{Y} \end{pmatrix}, \tag{12}$$

2.3. Least squares support vector machine for binary classification

Considering a training set of l pairs of samples $\{\mathbf{x}_i, y_i\}_{i=1}^{l}$ for binary classification, where $\mathbf{x}_i \in R^n$ are the input data and $y_i \in \{-1, +1\}$ are the corresponding class labels, LS-SVM for classification problem is also a QP problem based on the equality constraints and quadratic loss function, and can be described in the following [18,19]:

$$\min_{\mathbf{w},b,\xi} J(\mathbf{w}, b, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2, \tag{13}$$

s. t.

$$y_i(\mathbf{w}^T\varphi(\mathbf{x}_i) + b) = 1 - \xi_i, \quad i = 1, \cdots, l, \tag{14}$$

2.4. Weighted least squares support vector machine for binary classification

WLS-SVM for binary classification is described in the following [46]:

$$\min_{\mathbf{w},b,\xi} J(\mathbf{w}, b, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{l}s_i\xi_i^2, \tag{15}$$

s. t.

$$y_i(\mathbf{w}^T\varphi(\mathbf{x}_i) + b) = 1 - \xi_i, \quad i = 1, \cdots, l. \tag{16}$$

Multiplying y_i in both sides of (16) yields

$$y_i - (\mathbf{w}^T\varphi(\mathbf{x}_i) + b) = y_i\xi_i, \quad i = 1, \cdots, l. \tag{17}$$

Download English Version:

https://daneshyari.com/en/article/406666

Download Persian Version:

https://daneshyari.com/article/406666

Daneshyari.com