ARTICLE IN PRESS YACHA:1132

Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Kernel-based sparse regression with the correntropy-induced loss

Hong Chen [a,*], Yulong Wang [b]

[a] Institute of Applied Mathematics, College of Science, Huazhong Agricultural University, Wuhan 430070, China
[b] Faculty of Science and Technology, University of Macau, Macau 999078, China

ARTICLE INFO

Article history:
Received 20 August 2015
Received in revised form 26 March 2016
Accepted 22 April 2016
Available online xxxx
Communicated by Dominique Picard

MSC:
68T05
62J02

Keywords:
Learning theory
Kernel-based regression
Correntropy-induced loss
Sparsity
Learning rate

ABSTRACT

The correntropy-induced loss (C-loss) has been employed in learning algorithms to improve their robustness to non-Gaussian noise and outliers recently. Despite its success on robust learning, only little work has been done to study the generalization performance of regularized regression with the C-loss. To enrich this theme, this paper investigates a kernel-based regression algorithm with the C-loss and ℓ_1-regularizer in data dependent hypothesis spaces. The asymptotic learning rate is established for the proposed algorithm in terms of novel error decomposition and capacity-based analysis technique. The sparsity characterization of the derived predictor is studied theoretically. Empirical evaluations demonstrate its advantages over the related approaches.

1. Introduction

The kernel-based regression under Tikhonov regularization framework is a fundamental study field in the machine learning community. From the aspect of kernelized dictionary learning, many least square regression methods have been proposed with different regularization versions, e.g., the kernel norm regularization in [7,25], ℓ_1-regularization in [22,29], ℓ_2-regularization in [35,30,23], and the elastic net regularization in [9,11]. Following this line, learning theory analysis has been established to provide their probabilistic generalization bounds.

Despite progress on algorithmic design and analysis, these works are limited to the least square loss which is induced by the mean squared error (MSE) criterion. The MSE criterion is optimal when the error distribution is Gaussian, and is sensitive to non-Gaussian noise. However, in many applications, samples

* Corresponding author.
E-mail addresses: chenh@mail.hzau.edu.cn (H. Chen), wangyulong6251@gmail.com (Y. Wang).

http://dx.doi.org/10.1016/j.acha.2016.04.004
1063-5203/

Please cite this article in press as: H. Chen, Y. Wang, Kernel-based sparse regression with the correntropy-induced loss, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.04.004

may be contaminated by the outliers or non-Gaussian noise. This motives the robust regression models by replacing the least square loss with other robust losses, e.g., the Huber loss in [17], the pinball loss in [18,26], and the entropy-based losses in [21,28,10].

Among these losses, the entropy-based losses can be mainly divided into two classes: the loss associated with the minimum error entropy (MEE) criterion and the loss induced by maximum correntropy criterion (MCC). The MEE criterion and MCC have been well studied in information theoretical learning (ITL) [21], and have been used widely for signal processing [27,19] and machine learning [12,13,28]. For the problem involving non-Gaussian noise, MEE might perform well since all order moments of the error variable are taken into account by the Rényi's entropy [15,8]. However, this entropy conduction is done over all pairs of error variables, which makes the computation heavy for large training samples [27,10]. To improve algorithm efficiency, MCC is used for robust prediction since it does not require the computing effort of MEE, and also can well deal with Gaussian and non-Gaussian noise. Therefore, this paper considers the regression problem with the correntropy-induced loss (C-loss).

Recently, the properties of C-loss have been investigated in [28] for classification and its effectiveness has been demonstrated in data experiments via neural networks. The C-loss also has been used in [10] for robust regression associated with reproducing kernel Hilbert spaces (RKHS), where learning theory analysis and empirical evaluation are provided. Inspired from this study, we propose a sparse regularized regression framework with the C-loss and general kernel function, which is not necessary to be a Mercer kernel. For the kernelized sparse dictionary learning, only part samples are required to construct the predictor, which much improves the algorithm efficiency [22,3,4]. Although the sparse regularization results in the additional difficulty in error analysis, we overcome it in terms of a novel error decomposition and the characteristics of data dependent hypothesis spaces. Learning theory analysis is presented to illustrate the generalization performance of the robust kernel-based sparse regression algorithm (RKSR).

As a summary, the key features of this paper are listed as below:

- The proposed model brings MCC and kernelized dictionary learning together, which incorporates the C-loss and the ℓ_1-regularizer in a coefficient-based regularization scheme. Moreover, the sparse representation of this model is characterized theoretically. To the best of our knowledge, this sparse regularized model with C-loss has not been investigated previously in machine learning literatures.
- The generalization error bound is established in terms of novel error decomposition and concentration inequality associated with the empirical covering numbers. In particular, the learning rate with polynomial decay $O(m^{-\frac{1}{2}})$ can be reached, which is consistent with the related results in [24] for sparse least square regression and does not need the assumption on marginal distribution.
- The proposed algorithm is feasible to implement via the half-quadratic optimization technique [20,12,13]. The empirical studies on simulated and benchmark datasets demonstrate the advantages over the related approaches.

The rest of this paper is organized as follows. In Section 2, we introduce the RKSR with the C-loss and data dependent hypothesis space. The main theoretical analysis on learning rate and sparsity characterization are established in Section 3, and the analysis on computational complexity is presented in Section 4. The proof of generalization bound is given in Section 5. In Section 6, the effectiveness of RKSR is verified on data experiments. Finally, we conclude this paper in Section 7.

2. Robust kernel-based sparse regression

Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact metric space and $\mathcal{Y} \subset [-M, M]$ with some constant $M > 0$. Denote ρ as the probability distribution on $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ and $\rho(\cdot|x)$ as the conditional distribution for given $x \in \mathcal{X}$. Assume that the samples $\mathbf{z} = \{z_i\}_{i=1}^m = \{(x_i, y_i)\}_{i=1}^m$ are drawn independently according to ρ. Given $\mathbf{z}$, the regression problem aims to find a good approximation of the regression function

Please cite this article in press as: H. Chen, Y. Wang, Kernel-based sparse regression with the correntropy-induced loss, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.04.004