Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Q1

Penalized principal logistic regression for sparse sufficient dimension reduction

Q2 Seung Jun Shin *, Andreas Artemiou

Korea University, South Korea
Cardiff University, United Kingdom

A R T I C L E I N F O

Article history:
Received 3 June 2016
Received in revised form 20 September 2016
Accepted 5 December 2016
Available online xxxx

Keywords:
Max-SCAD penalty
Principal logistic regression
Sparse sufficient dimension reduction
Sufficient dimension reduction

A B S T R A C T

Sufficient dimension reduction (SDR) is a successive tool for reducing the dimensionality of predictors by finding the central subspace, a minimal subspace of predictors that preserves all the regression information. When predictor dimension is large, it is often assumed that only a small number of predictors is informative. In this regard, sparse SDR is desired to achieve variable selection and dimension reduction simultaneously. We propose a principal logistic regression (PLR) as a new SDR tool and further develop its penalized version for sparse SDR. Asymptotic analysis shows that the penalized PLR enjoys the oracle property. Numerical investigation supports the advantageous performance of the proposed methods.

1. Introduction

It is often of primary interest to identify the relationship between the univariate response Y and the p-dimensional predictor $\mathbf{X} \in \mathbb{R}^p$. Sufficient dimension reduction (SDR) efficiently reduces the dimensionality of $\mathbf{X}$ by finding a lower dimensional subspace of span($\mathbf{X}$) while preserving regression information in $\mathbf{X}$. Specifically, SDR seeks a matrix $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_d) \in \mathbb{R}^{p\times d}$ that satisfies

$$Y \perp \mathbf{X} | \mathbf{B}^\top \mathbf{X}, \tag{1}$$

where $\perp$ denotes statistical independence. Compared to conventional parametric models, (1) is less stringent since it does not assume any specific link functions between Y and $\mathbf{X}$. The space spanned by $\mathbf{B}$ satisfying (1) is called the dimension reduction subspace (DRS). The central subspace, denoted by $\mathcal{S}_{Y|\mathbf{X}}$ is defined as the intersection of all DRSes, and hence it is the lowest dimensional DRS. Cook (1996) showed that $\mathcal{S}_{Y|\mathbf{X}}$ uniquely exists under mild conditions. In SDR, it is assumed that $\mathcal{S}_{Y|\mathbf{X}} = \text{span}(\mathbf{B})$ to make $\mathbf{B}$ an identifiable target. The dimension d of $\mathcal{S}_{Y|\mathbf{X}}$ is referred to as the structural dimension, another important quantity to be inferred from the data.

Since the seminal paper on sliced inverse regression (SIR, Li, 1991), there have been various methods developed to estimate $\mathcal{S}_{Y|\mathbf{X}}$, which include but are not limited to sliced averaged variance estimation (SAVE, Cook and Weisberg, 1991), directional regression (DR, Li and Wang, 2007), sliced regression (Wang and Xia, 2008), contour regression (Li et al., 2005), and principal support vector machine (PSVM, Li et al., 2011).

* Correspondence to: Department of statistics, Korea university, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. Fax: +82 2 924 9895.
E-mail address: sjshin@korea.ac.kr (S.J. Shin).

http://dx.doi.org/10.1016/j.csda.2016.12.003
0167-9473/© 2016 Elsevier B.V. All rights reserved.

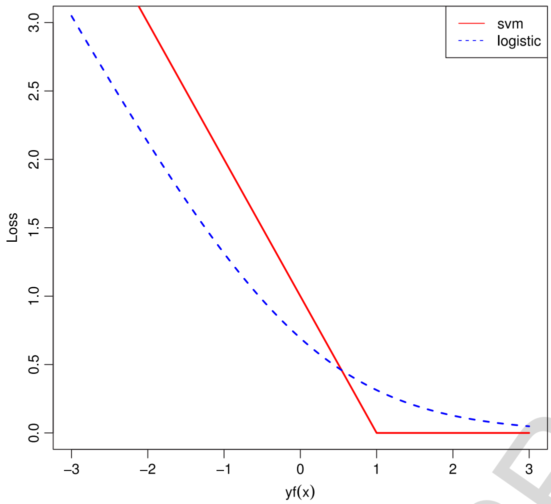

Fig. 1. SVM (hinge) loss versus logistic (binomial log-likelihood) loss: The (dashed blue) logistic loss is a smooth, convex, and continuously differentiable function while the (solid red) SVM hinge loss is not. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Among many others, PSVM is a recently developed SDR method and brings new insight by connecting SDR to penalized machine learning methods such as the support vector machine (SVM). The idea of PSVM is simple as follows. First, dichotomize the continuous response Y by introducing a pseudo response $\tilde{Y} = 1$ if Y is greater than a given cutoff value r, and -1 otherwise. A sequence of linear SVMs is then repeatedly trained for $(\tilde{Y}_r, \mathbf{X})$ for different values of r. Li et al. (2011) showed that normals of the optimal hyperplanes from these linear SVMs lie on $\mathcal{S}_{Y|\mathbf{X}}$. Finally, $\mathcal{S}_{Y|\mathbf{X}}$ can be recovered by the spectral decomposition of these normals. PSVM is known to perform better than classical SDR methods such as SIR, and it tackles both linear and nonlinear SDR in a unified framework via kernel trick, as SVM does.

In this article, we propose a principal logistic regression (PLR) as an alternative to PSVM. Namely, we apply the logistic regression to $(\tilde{Y}, \mathbf{X})$ instead of SVM. The advantages of the logistic regression over SVM are obvious since its loss function is smooth and strictly convex (see Fig. 1). PLR not only entails simpler asymptotic results under less stringent conditions but also is computationally stable. It is important to note that PLR is not a parametric method for SDR since we replace the loss in population level and hence the target of estimation completely changes.

Sparse SDR that seeks a sparse representation of the basis of $\mathcal{S}_{Y|\mathbf{X}}$ is often desired to achieve the dimension reduction and variable selection simultaneously. Sparse SDR facilitates the interpretation of the results and improves the estimation accuracy by eliminating negligible uncertainties from the predictors with weak signals (Li, 2007). Toward sparse SDR, several methods have been proposed. See, for example, Li (2007), Bondell and Li (2009), Chun and Keleş (2010) and Wu and Li (2011).

Sparse SDR assumes a unique partition $\mathbf{X}^\top = (\mathbf{X}_+^\top, \mathbf{X}_-^\top)$ that satisfies

$$Y \perp \mathbf{X}_- | \mathbf{X}_+, \tag{2}$$

where $\mathbf{X}_+ \in \mathbb{R}^q$ and $\mathbf{X}_- \in \mathbb{R}^{p-q}$ for some $q(\ll p)$ (Cook et al., 2004; Bondell and Li, 2009). We call $\mathbf{X}_+$ and $\mathbf{X}_-$ relevant and irrelevant variables, respectively. Without loss of generality we assume that the first q predictors are the relevant ones throughout this article. Under (1) and (2), the last $p-q$ rows of $\mathbf{B}$ are all zeros, which makes $\mathbf{B}$ sparse and has an identical sparsity structure across different columns. That is, the last $p-q$ elements of $\mathbf{B}$ are zeros regardless of the cutoff values. In order to preserve such sparsity structure, we employ a max-SCAD penalty. The SCAD penalty (Fan and Li, 2001) is known to enjoy the oracle property, but its computation is more challenging due to its nonconvexity. However, the logistic loss can minimize the additional computational burden thanks to its smoothness. We establish the oracle property of the max-SCAD penalized PLR and develop an efficient algorithm for its sample estimation.

The rest of the article is organized as follows. In Section 2 we propose PLR and describe related details including its sample estimation, asymptotic properties, and the structural dimension estimation. The penalized PLR is developed in Section 3 in which we establish its oracle property and develop an efficient algorithm for the sample estimation. In Section 4, simulation studies are carried out to investigate finite sample performances of both PLR and the penalized PLR, and real data analysis results are given in Section 5. Final discussions follow in Section 6. All the technical proofs are relegated to the online Supplementary Materials (see Appendix A).

Please cite this article in press as: Shin, S.J., Artemiou, A., Penalized principal logistic regression for sparse sufficient dimension reduction. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.12.003