Journal of Multivariate Analysis 134 (2015) 71–81

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Robust inverse regression for dimension reduction

Yuexiao Dong [a], Zhou Yu [b], Liping Zhu [c,d,*]

[a] Department of Statistics, Temple University, Philadelphia, PA, 19122, United States
[b] School of Finance and Statistics, East China Normal University, Shanghai, 200241, China
[c] School of Statistics and Management, Shanghai University of Finance and Economics (SUFE), Shanghai, 200433, China
[d] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai, 200433, China

ARTICLE INFO

Article history:
Received 24 July 2013
Available online 7 November 2014

AMS subject classifications:
62H12
62H30
62F35
62G35

Keywords:
Central space
Ellipticity
Multivariate median
Sliced inverse regression

ABSTRACT

Classical sufficient dimension reduction methods are sensitive to outliers present in predictors, and may not perform well when the distribution of the predictors is heavy-tailed. In this paper, we propose two robust inverse regression methods which are insensitive to data contamination: weighted inverse regression estimation and sliced inverse median estimation. Both weighted inverse regression estimation and sliced inverse median estimation produce unbiased estimates of the central space when the predictors follow an elliptically contoured distribution. Our proposals are compared with existing robust dimension reduction procedures through comprehensive simulation studies and an application to the New Zealand mussel data. It is demonstrated that our methods have better overall performances than existing robust procedures in the presence of potential outliers and/or inliers.

1. Introduction

In light of recent development of science and technology, high-dimensional data are collected at an unprecedented speed in various fields such as biology, economics and finance, etc. Analysis of high-dimensional data calls for new statistical theories and methodologies. A natural way to analyze high-dimensional data is to first reduce the dimensionality of the original data without losing vital information, then carry out ensuing statistical analysis based on the reduced data. Suppose we consider the regression of a univariate response Y onto a p-dimensional predictor vector $\mathbf{X}$. Sufficient dimension reduction [5] aims at finding low-dimensional linear combinations of the predictors which contain all the regression information, that is,

$$Y \perp\!\!\!\perp \mathbf{X} \mid \mathbf{B}^{\mathrm{T}}\mathbf{X}, \tag{1}$$

where $\perp\!\!\!\perp$ stands for statistical independence and $\mathbf{B} \in \mathbb{R}^{p\times d}$ with $d < p$. The conditional independence model (1) implies that it suffices to use the reduced predictors $\mathbf{B}^{\mathrm{T}}\mathbf{X}$ to infer about how the conditional distribution of $Y|\mathbf{X}$ changes with the values of $\mathbf{X}$. For any matrix $\mathbf{B}$ satisfying (1), its column space is a dimension reduction space. When it exists, the intersection of all dimension reduction spaces is called the central space, and denoted by $\mathcal{S}_{Y|\mathbf{X}}$ [4]. The dimension of $\mathcal{S}_{Y|\mathbf{X}}$, denoted by d, is called the structural dimension of the central space.

Sliced inverse regression [19] is an early attempt to estimate $\mathcal{S}_{Y|\mathbf{X}}$. It divides the observations into several slices according to the values of the response, then synthesizes the intraslice means of the predictors to estimate $\mathcal{S}_{Y|\mathbf{X}}$. Many proposals based

* Corresponding author at: School of Statistics and Management, Shanghai University of Finance and Economics (SUFE), Shanghai, 200433, China.
E-mail address: zhu.liping@mail.shufe.edu.cn (L. Zhu).

http://dx.doi.org/10.1016/j.jmva.2014.10.005
0047-259X/

on inverse conditional moments are developed thereafter, which include sliced average variance estimation [9], canonical correlation [15], contour regression [22], directional regression [21], and cumulative slicing estimation [34], etc. These methods relieve the curse of dimensionality [1] because they do not involve high-dimensional smoothing in estimating $\mathcal{S}_{Y|\mathbf{X}}$. All these methods require that the predictor vector $\mathbf{X}$ satisfies the linearity assumption, which is fulfilled when $\mathbf{X}$ has an elliptically contoured distribution with finite moments. When the predictor distribution is not elliptically contoured, Cook and Nachtsheim [7] suggested a reweighting technique to achieve the desired distribution. Recently, the idea of the central solution space [20,12] was proposed for non-elliptically distributed predictors.

We will focus on elliptically contoured predictors in this paper. Suppose throughout this paper that the density of $\mathbf{X}$ at $\mathbf{x} \in \mathbb{R}^p$ has the form

$$f(\mathbf{x}) = |\boldsymbol{\Gamma}|^{-1/2} g\left(\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Gamma}}\right), \quad \text{for some function } g(\cdot), \tag{2}$$

where $\|\mathbf{x}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Gamma}} = (\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, $\boldsymbol{\mu} \in \mathbb{R}^p$ is a $p \times 1$ vector, and $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times p}$ is a positive definite matrix. When $\mathbf{X}$ is elliptically contoured with infinite moments, or when the elliptically contoured predictor $\mathbf{X}$ is subject to data contamination, classical moment-based sufficient dimension reduction methods such as sliced inverse regression may fail.

Data contamination may be in the form of either outliers or inliers. The outliers are data values which lie in the tail of the statistical distribution of a set of data values. The inliers, on the other hand, are data values that lie in the interior of a statistical distribution and are in error. The inliers are often-times difficult to be distinguished from normal data values. An example of an inlier might be a value in wrong units, say degrees Celsius instead of degrees Fahrenheit. Inliers can arise due to systematic error and certain types of respondent and processing error. Existing robust sufficient dimension reduction procedures in the literature focus on outlier detection and correction, which include contour projection [29,23] and weighted canonical correlation [31]. Contour projection first normalizes $\mathbf{X}$ to $\tilde{\mathbf{X}} = (\mathbf{X}-\boldsymbol{\mu})/\|\mathbf{X}-\boldsymbol{\mu}\|_{\boldsymbol{\Gamma}}$, and then identifies $\mathcal{S}_{Y|\mathbf{X}}$ through $\mathcal{S}_{Y|\tilde{\mathbf{X}}}$. Weighted canonical correlation uses weight function $1/(1+\|\mathbf{X}-\boldsymbol{\mu}\|^2_{\boldsymbol{\Gamma}})$ to downweight potential outliers, and then performs classical canonical correlation [15] with the transformed predictors. Contour projection and weighted canonical correlation are not sensitive to outliers, but they may inadvertently overweight inliers. We will see that in the presence of inliers, both methods may be outperformed by the classical sliced inverse regression, which is only sensitive to outliers and not sensitive to inliers.

In this paper we propose two robust inverse regression methods: weighted inverse regression estimation and sliced inverse median estimation. Both methods (i) perform comparably to the classical sufficient dimension reduction procedures when the predictors have finite moments with no outliers or inliers, and (ii) keep good performances in the presence of outliers and inliers, and/or when the predictors have infinite moments. Due to the invariance law of central space [5], we have $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Gamma}^{-1/2}\mathcal{S}_{Y|\mathbf{Z}}$, where $\mathbf{Z} = \boldsymbol{\Gamma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})$ is the standardized predictor. Without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Gamma} = \mathbf{I}_p$, and work with $\mathbf{Z}$ and $\mathcal{S}_{Y|\mathbf{Z}}$ for our population level development. Furthermore, denote $J_1, \ldots, J_H$ as a partition for the range of Y. For categorical response, such a partition is obvious and H denotes the number of categories. For continuous response, the partition could be taken as dividing the ordered response into H equally sized slices. For any matrix $\mathbf{M}$, denote its column space as $\mathrm{Span}(\mathbf{M})$.

2. Weighted inverse regression estimation

We propose a weighted version of sliced inverse regression in this section. From (2), the standardized predictor $\mathbf{Z}$ has a spherically contoured distribution, and the density of $\mathbf{Z}$ at $\mathbf{z} \in \mathbb{R}^p$ only depends on $\|\mathbf{z}\| = (\mathbf{z}^{\mathrm{T}}\mathbf{z})^{1/2}$. For $\mathbf{A} \in \mathbb{R}^{p\times d}$ satisfying $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I}_d$ and $Y \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{A}^{\mathrm{T}}\mathbf{Z}$, let $\mathbf{P}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ be the projection matrix onto the column space of $\mathbf{A}$. Assuming its existence, it is easy to see that the conditional expectation $E(\mathbf{Z}|\mathbf{A}^{\mathrm{T}}\mathbf{Z})$ equals $\mathbf{P}_{\mathbf{A}}\mathbf{Z}$. Following Li [19], we have

$$E(\mathbf{Z}|Y) = E\{E(\mathbf{Z}|Y, \mathbf{A}^{\mathrm{T}}\mathbf{Z})|Y\} = E\{E(\mathbf{Z}|\mathbf{A}^{\mathrm{T}}\mathbf{Z})|Y\} = \mathbf{P}_{\mathbf{A}}E(\mathbf{Z}|Y) \subseteq \mathcal{S}_{Y|\mathbf{Z}}.$$

Thus we can use $E(\mathbf{Z}|Y)$ to recover $\mathcal{S}_{Y|\mathbf{Z}}$ as long as the moments involved exist. However, $E(\mathbf{Z}|Y)$ may not exist for heavy-tailed distribution such as multivariate Cauchy. This observation motivates us to propose weighted inverse regression estimation.

Recall that $J_1, \ldots, J_H$ is a partition for the range of Y. Let $\kappa(\cdot)$ be a weighting function such that $E\{\kappa(\|\mathbf{Z}\|)\mathbf{Z}|Y \in J_h\}$ exists for $h = 1, \ldots, H$. Denote $p_h = E(I_h(Y))$, where $I_h(Y)$ is the indicator function of $Y \in J_h$. Let $\boldsymbol{\mu}_{\kappa,h} = E\{\kappa(\|\mathbf{Z}\|)\mathbf{Z}|Y \in J_h\}$ and define the following kernel matrix

$$\mathbf{K}_{\mathrm{WIRE}} \equiv \sum_{h=1}^{H} p_h \boldsymbol{\mu}_{\kappa,h}\boldsymbol{\mu}_{\kappa,h}^{\mathrm{T}}. \tag{3}$$

The next result establishes the validity of estimating $\mathcal{S}_{Y|\mathbf{Z}}$ via the column space of $\mathbf{K}_{\mathrm{WIRE}}$.

Theorem 1. *Suppose $\mathbf{Z}$ has a spherically contoured distribution. Then if all the moments involved exist,* $\mathrm{span}(\mathbf{K}_{\mathrm{WIRE}}) \subseteq \mathcal{S}_{Y|\mathbf{Z}}$.

In the general case when $\mathbf{X}$ has an elliptically contoured distribution as in (2), Theorem 1 implies that $\boldsymbol{\Gamma}^{-1/2}\mathrm{span}(\mathbf{K}_{\mathrm{WIRE}}) \subseteq \mathcal{S}_{Y|\mathbf{X}}$.

Contour projection of sliced inverse regression [29] is a special case of our proposal when we set $\kappa(\|\mathbf{Z}\|) = 1/\|\mathbf{Z}\|$. The weighted canonical correlation method [31] is a weighted version of canonical correlation [15], and the corresponding weight is proportional to $\kappa(\|\mathbf{Z}\|) = 1/(1+\|\mathbf{Z}\|^2)$. Both methods mitigate the effect of potential outliers, as the weight is small for large $\|\mathbf{Z}\|$. However, contour projection and weighted canonical correlation may end up giving unduly large weights to