Statistics and Probability Letters 139 (2018) 90–94

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On a new class of sufficient dimension reduction estimators

Yuexiao Dong *, Yongxu Zhang

Department of Statistical Science, Temple University, Philadelphia, PA 19122, United States

ARTICLE INFO

Article history:
Received 28 October 2017
Received in revised form 12 March 2018
Accepted 28 March 2018
Available online 9 April 2018

Keywords:
Linear conditional mean
Ordinary least squares
Sliced inverse regression

ABSTRACT

OLS and SIR are two popular sufficient dimension reduction estimators. OLS can recover at most one direction, and SIR shares this limitation when the response is binary. To address such limitation, we propose slicing-assisted OLS and slicing-assisted SIR.

1. Introduction

As a popular tool for multivariate analysis, sufficient dimension reduction (SDR) (Li, 1991; Cook, 1998) aims to reduce the predictor dimension and to contain the relevant regression information between the response and the predictor. For univariate response $Y \in \mathbb{R}$ and multivariate predictor $\mathbf{X} \in \mathbb{R}^p$, one goal of SDR is to find $\mathbf{B} \in \mathbb{R}^{p\times d}$ with $d < p$, such that

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

where "$\perp\!\!\!\perp$" means statistical independency. If $\mathbf{B}$ satisfies (1), then the column space of $\mathbf{B}$ is called a dimension reduction space. Under mild assumptions (Yin et al., 2008), the intersection of all dimension reduction spaces is itself a dimension reduction space, and is known as the central space. We denote the central space by $\mathcal{S}_{Y|\mathbf{X}}$. When the focus is on the regression mean, we consider

$$Y \perp\!\!\!\perp E(Y|\mathbf{X}) \mid \mathbf{B}^{\mathrm{T}}\mathbf{X}. \tag{2}$$

If $\mathbf{B}$ satisfies (2), then the column space of $\mathbf{B}$ is called a mean dimension reduction space. The intersection of all mean dimension reduction spaces is called the central mean space (Cook and Li, 2002), and is denoted by $\mathcal{S}_{E(Y|\mathbf{X})}$. Let $\mathbf{B} \in \mathbb{R}^{p\times d}$ be the basis of $\mathcal{S}_{E(Y|\mathbf{X})}$ (or $\mathcal{S}_{Y|\mathbf{X}}$), and d is known as the structural dimension of the central mean space (or the central space).

Ordinary least squares (OLS) (Li and Duan, 1989) is a classical SDR estimator for the central mean space. An obvious limitation of OLS is that it can estimate at most one direction in the central mean space. In this short note, we propose slicing-assisted OLS (SOLS) to address this limitation. Our key idea is to slice the support of $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$, where $\boldsymbol{\beta}_0$ denotes the original OLS estimator. SOLS improves the original OLS as it can recover more than one direction in the central mean space. Sliced inverse regression (SIR) (Li, 1991), on the other hand, is a popular SDR method for the central space. In the case of binary response, it is well-known that SIR can only recover at most one direction in the central space. Following similar development as the SOLS, we propose slicing-assisted SIR (SSIR), which can recover multiple directions in the central space with binary response.

* Corresponding author.
E-mail address: ydong@temple.edu (Y. Dong).

https://doi.org/10.1016/j.spl.2018.03.019
0167-7152/

The rest of the paper is organized as follows. Section 2 introduces the slicing-assisted OLS. Section 3 studies the slicing-assisted SIR with binary response. Simulation studies are presented in Section 4. All the proofs are relegated to the Appendix. We assume the structural dimension d is known throughout the paper.

2. Slicing-assisted OLS

Denote $E(\mathbf{X})=\boldsymbol{\mu}$ and $\mathrm{Var}(\mathbf{X})=\boldsymbol{\Sigma}$. The classical OLS estimator is $\boldsymbol{\beta}_0=\boldsymbol{\Sigma}^{-1}E\{(\mathbf{X}-\boldsymbol{\mu})Y\}$. Suppose $\mathrm{Span}(\mathbf{B})=\mathcal{S}_{E(Y|\mathbf{X})}$, where $\mathrm{Span}(\cdot)$ denotes the column space. The following assumption is common in the sufficient dimension reduction literature:

$$E(\mathbf{X}|\mathbf{B}^{\mathrm{T}}\mathbf{X}) \text{ is a linear function of } \mathbf{B}^{\mathrm{T}}\mathbf{X}. \tag{3}$$

Assumption (3) is referred to as the linear conditional mean (LCM) assumption. Under LCM, it can be shown that $\boldsymbol{\beta}_0\in\mathcal{S}_{E(Y|\mathbf{X})}$, which means OLS can be used to recover the central mean space. OLS can recover at most one dimension in the central mean space, and this motivates us to propose slicing-assisted OLS. Our proposal is based on the following key observation.

Theorem 2.1. *Assume (3) holds for* $\mathbf{B}$*, where* $\mathrm{Span}(\mathbf{B})=\mathcal{S}_{E(Y|\mathbf{X})}$*. Then* $\boldsymbol{\Sigma}^{-1}E\{(\mathbf{X}-\boldsymbol{\mu})Y|{\boldsymbol{\beta}_0}^T\mathbf{X}\}\in\mathcal{S}_{E(Y|\mathbf{X})}$.

Recall that the original OLS estimator is $\boldsymbol{\beta}_0=\boldsymbol{\Sigma}^{-1}E\{(\mathbf{X}-\boldsymbol{\mu})Y\}$. Our proposal is to replace the unconditional expectation in OLS with the conditional expectation on $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$.

Let $J_1,\dots,J_H$ be a partition for the support of $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$. For $h=1,\dots,H$, denote $I(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\in J_h)$ as the indicator function of $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}$ belonging to J_h. Let $\pi_h=E\{I(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\in J_h)\}$, $\mathbf{U}_h=E\{(\mathbf{X}-\boldsymbol{\mu})YI(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{X}\in J_h)\}$, and define

$$\mathbf{M}_{\mathrm{sols}}=\boldsymbol{\Sigma}^{-1}\left(\sum_{h=1}^{H}\pi_h^{-1}\mathbf{U}_h\mathbf{U}_h^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{-1}. \tag{4}$$

The next result follows directly from Theorem 2.1.

Corollary 2.1. *Assume (3) holds for* $\mathbf{B}$*, where* $\mathrm{Span}(\mathbf{B})=\mathcal{S}_{E(Y|\mathbf{X})}$*. Then* $\mathrm{Span}(\mathbf{M}_{sols})\subseteq\mathcal{S}_{E(Y|\mathbf{X})}$.

Corollary 2.1 suggests that the sample version of $\mathbf{M}_{\mathrm{sols}}$ in (4) can be used to recover multiple directions in the central mean space. Let $\{(\mathbf{X}_i,Y_i), i=1,\dots,n\}$ be an i.i.d. sample of $(\mathbf{X},Y)$. We conclude this section with the sample level SOLS algorithm.

1. Let $\bar{\mathbf{X}}=n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$, $\hat{\mathbf{X}}_i=\mathbf{X}_i-\bar{\mathbf{X}}$, $\hat{\boldsymbol{\Sigma}}=n^{-1}\sum_{i=1}^{n}\hat{\mathbf{X}}_i\hat{\mathbf{X}}_i^{\mathrm{T}}$, $E_n(\mathbf{X}Y)=n^{-1}\sum_{i=1}^{n}\hat{\mathbf{X}}_iY_i$, and $\hat{\boldsymbol{\beta}}_0=\hat{\boldsymbol{\Sigma}}^{-1}E_n(\mathbf{X}Y)$.
2. Calculate $\hat{\mathbf{M}}_{\mathrm{sols}}=\hat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{h=1}^{H}\hat{\pi}_h^{-1}\hat{\mathbf{U}}_h\hat{\mathbf{U}}_h^{\mathrm{T}}\right)\hat{\boldsymbol{\Sigma}}^{-1}$, where for $h=1,\dots,H$, $\hat{\pi}_h=n^{-1}\sum_{i=1}^{n}I(\hat{\boldsymbol{\beta}}_0^{\mathrm{T}}\hat{\mathbf{X}}_i\in J_h)$ and $\hat{\mathbf{U}}_h=n^{-1}\sum_{i=1}^{n}\hat{\mathbf{X}}_iY_iI(\hat{\boldsymbol{\beta}}_0^{\mathrm{T}}\hat{\mathbf{X}}_i\in J_h)$.
3. Perform eigenvalue decomposition of $\hat{\mathbf{M}}_{\mathrm{sols}}$. Let $\hat{\boldsymbol{\beta}}_1^{\mathrm{sols}},\dots,\hat{\boldsymbol{\beta}}_d^{\mathrm{sols}}$ be the eigenvectors corresponding to the d largest eigenvalues of $\hat{\mathbf{M}}_{\mathrm{sols}}$. The final SOLS estimator of $\mathcal{S}_{E(Y|\mathbf{X})}$ is $\mathrm{Span}(\hat{\boldsymbol{\beta}}_1^{\mathrm{sols}},\dots,\hat{\boldsymbol{\beta}}_d^{\mathrm{sols}})$.

3. Slicing-assisted SIR with binary response

We discuss estimators of the central space in this section. Suppose the LCM assumption (3) holds for $\mathbf{B}$, where $\mathrm{Span}(\mathbf{B})=\mathcal{S}_{Y|\mathbf{X}}$. The classical SIR is based on the fact that $\boldsymbol{\Sigma}^{-1}E(\mathbf{X}-\boldsymbol{\mu}|Y)\in\mathcal{S}_{Y|\mathbf{X}}$.

In the case of binary response, denote the two categories as $Y=0$ or $Y=1$. Let $\boldsymbol{\xi}_0=\boldsymbol{\Sigma}^{-1}E(\mathbf{X}-\boldsymbol{\mu}|Y=0)$ and $\boldsymbol{\xi}_1=\boldsymbol{\Sigma}^{-1}E(\mathbf{X}-\boldsymbol{\mu}|Y=1)$. Note that $p_0\boldsymbol{\xi}_0+p_1\boldsymbol{\xi}_1=\mathbf{0}$, where $p_0=\Pr(Y=0)$ and $p_1=\Pr(Y=1)$. Thus SIR recovers only one unique direction in the central space. This motivates us to propose slicing-assisted SIR. The following result is parallel to Theorem 2.1.

Theorem 3.1. *Assume (3) holds for* $\mathbf{B}$*, where* $\mathrm{Span}(\mathbf{B})=\mathcal{S}_{Y|\mathbf{X}}$*. Then* $\boldsymbol{\Sigma}^{-1}E(\mathbf{X}-\boldsymbol{\mu}|Y,{\boldsymbol{\xi}_0}^T\mathbf{X})\in\mathcal{S}_{Y|\mathbf{X}}$.

Recall that the original SIR estimator is $\boldsymbol{\Sigma}^{-1}E(\mathbf{X}-\boldsymbol{\mu}|Y)$. Our proposal is to replace the conditional expectation on Y with the conditional expectation on both Y and $\boldsymbol{\xi}_0^{\mathrm{T}}\mathbf{X}$. Unlike SIR that only requires slicing the support of Y, the new estimator requires double slicing through both the support of Y and the support of $\boldsymbol{\xi}_0^{\mathrm{T}}\mathbf{X}$.

Let $K_1,\dots,K_H$ be a partition for the support of $\boldsymbol{\xi}_0^{\mathrm{T}}\mathbf{X}$. For $\ell=1,2$ and $h=1,\dots,H$, denote $I_{\ell,h}=I(Y=\ell,\boldsymbol{\xi}_0^{\mathrm{T}}\mathbf{X}\in K_h)$ as the indicator function of $\boldsymbol{\xi}_0^{\mathrm{T}}\mathbf{X}$ belonging to K_h and $Y=\ell$. Let $\tau_{\ell,h}=E(I_{\ell,h})$, $\mathbf{V}_{\ell,h}=E\{(\mathbf{X}-\boldsymbol{\mu})I_{\ell,h}\}$, and define

$$\mathbf{M}_{\mathrm{ssir}}=\boldsymbol{\Sigma}^{-1}\left(\sum_{\ell=0}^{1}\sum_{h=1}^{H}\tau_{\ell,h}^{-1}\mathbf{V}_{\ell,h}\mathbf{V}_{\ell,h}^{\mathrm{T}}\right)\boldsymbol{\Sigma}^{-1}. \tag{5}$$

The next result follows from Theorem 3.1.