Localized Sufficient Dimension Reduction [1]

Qin Wang

Department of Statistics
The University of Georgia
Athens, GA 30602

July 29, 2009

[1] Joint work with Dr. Xiangrong Yin (UGA), Dr. Bing Li and Dr. Zhihui Tang (PSU)

Outline

1. Introduction
2. Localized Sufficient Dimension Reduction
3. KNN Sliced Inverse Regression
4. Simulation Studies
5. Summary and Future Directions

Outline

1. Introduction
2. Localized Sufficient Dimension Reduction
3. KNN Sliced Inverse Regression
4. Simulation Studies
5. Summary and Future Directions

Challenge of High Dimensional Data

- **Easy** — Collect information
  - Biotech Data;
  - Financial Data;
  - Satellite image, etc.
- **Difficult** — Visualize and analyze data
  - Curse of dimensionality.

Data Structure

It is not uncommon that most of the data structure is concentrated on a low-dimensional subspace of the original high-dimensional space.

$$Y = f(\mathbf{X}, \epsilon) = f(x_1, x_2, \cdots, x_p, \epsilon) \tag{1.1}$$

can be transformed to

$$Y = f(\beta_1^T \mathbf{X}, \beta_2^T \mathbf{X}, \cdots, \beta_d^T \mathbf{X}, \epsilon) \tag{1.2}$$

where β's are unknown $p \times 1$ vectors and $d < p$.

Sufficient Dimension Reduction

$(\beta_1^T\mathbf{X}, \beta_2^T\mathbf{X}, \cdots, \beta_d^T\mathbf{X})$ captures all the information needed for the inference on $\mathbf{Y}$.

Sufficient dimension reduction — recover $(\beta_1, \beta_2, \cdots, \beta_d)$ without model fitting.

Central Space

- Dimension reduction is to find a d-dimensional projection subspace $\mathcal{S}$=Span$(\beta_1, \beta_2, \cdots, \beta_d)$ such that

$$Y \perp\!\!\!\perp \mathbf{X} | P_{\mathcal{S}} \mathbf{X} \tag{1.3}$$

$\mathcal{S}$ is then called a *dimension reduction subspace* (DRS) for $Y|\mathbf{X}$.

- *Central subspace* (CS) $\mathcal{S}_{Y|\mathbf{X}}$ — the intersection of all subspaces satisfying (1.3) if it also satisfies (1.3).
- Under mild conditions (Cook, 1998; Yin, Li and Cook, 2008), the CS exists and is unique.

Inverse Regression Methods

- A major class of dimension reduction methods;
- Inverse regression methods study the (inverse) conditional distribution of $\mathbf{X}|Y$, $F_{\mathbf{X}|Y}$;
- **Linearity condition.** For the directions $\mathbf{B} = (\beta_1, \beta_2, \cdots, \beta_d)$ in model (1.2) and any constant vector $\mathbf{b} \in \mathbb{R}^p$, there exist constants $c_0 \in \mathbb{R}^1$ and $\mathbf{c} \in \mathbb{R}^d$ depending on $\mathbf{b}$ such that $E(\mathbf{b}^T\mathbf{X}|\mathbf{B}^T\mathbf{X}) = c_0 + \mathbf{c}^T\mathbf{B}^T\mathbf{X}$.

Sliced Inverse Regression

- *Sliced inverse regression* (SIR; Li 1991) is the first and the most commonly used SDR method;
- SIR is investigating the trajectory of the inverse mean curve $E(\mathbf{X}|Y)$;

Let $\mathbf{Z} = \Sigma_{\mathbf{X}}^{-\frac{1}{2}}(\mathbf{X} - E(\mathbf{X}))$. Under linearity condition,

$$\mathcal{S}_{E(\mathbf{Z}|Y)} \subseteq \mathcal{S}_{Y|\mathbf{Z}} \tag{1.4}$$

Advantages

- Easy to implement;
  - Slicing Y;
  - Calculating $E(\mathbf{X}|Y)$;
  - Performing weighted PCA.
- Nice asymptotic properties.
  - Asymptotic normality of estimates;
  - Sequential test for d.

Disadvantages

- Probabilistic assumption;
- Non-exhaustiveness (coverage condition; Cook, 2004) ;
- Interpretability.

Question: Can we develop a new class of dimension reduction methods to overcome these disadvantages while keeping computation efficient?

Outline

1 Introduction

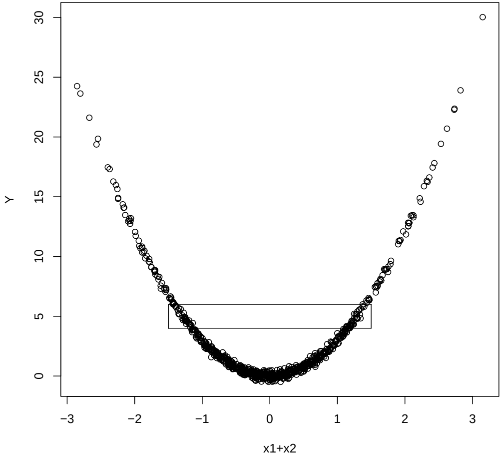
2 Localized Sufficient Dimension Reduction

3 KNN Sliced Inverse Regression

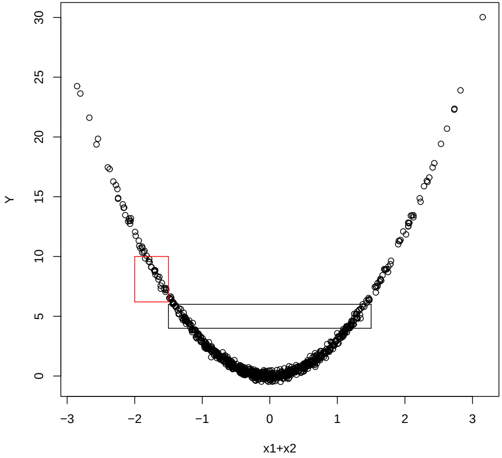
4 Simulation Studies

5 Summary and Future Directions

Motivation I



Motivation I



Motivation II

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

Localization of Dimension Reduction Subspace

Theorem

Suppose $\Omega_{\mathbf{X}}$ *is an open set in* $\mathbb{R}^p$. *Then*

1. $\mathcal{S}_{Y|\mathbf{X}} = \mathrm{span}\{\mathcal{S}_{Y_G|\mathbf{X}_G} : G \subseteq \Omega_{\mathbf{X}}\}$.
2. *Furthermore, there exist finite number of open sets, say* $G_1, \ldots, G_m$ *in* $\Omega_{\mathbf{X}}$, *such that* $\mathcal{S}_{Y|\mathbf{X}} = \mathrm{span}\{\mathcal{S}_{Y_{G_i}|\mathbf{X}_{G_i}} : i = 1, \ldots, m\}$.

Localization of Sliced Inverse Regression

Theorem

Suppose that, for a fixed $y \in \Omega_Y$, $g(y) > 0$, $h(y|\mathbf{x})$ is twice differentiable with respect to $\mathbf{x}$ on $\bar{G}$, and the second derivatives are bounded on $\bar{G}$. Then, for any $\mathbf{x} \in G$, and almost everywhere on Ω_Y,

$$h(y|\mathbf{x}) = h(y|\boldsymbol{\mu}_G + \mathbf{P}_{\beta_G}(\mathbf{x} - \boldsymbol{\mu}_G)) + O(\|G\|^2) \;\; as \;\; \|G\| \to 0.$$

Furthermore, as $\|G\| \to 0$, and almost everywhere on Ω_Y,

$$\boldsymbol{\Sigma}_G^{-1}[E(\mathbf{X}_G|y) - E(\mathbf{X}_G)] = \mathbf{P}_{\beta_G}\boldsymbol{\Sigma}_G^{-1}[E(\mathbf{X}_G|y) - E(\mathbf{X}_G)] + O(\|G\|).$$

Outline

1 Introduction

2 Localized Sufficient Dimension Reduction

3 KNN Sliced Inverse Regression

4 Simulation Studies

5 Summary and Future Directions

Estimation Procedure

(1) For an observation $(\mathbf{X}_s, Y_s)$, find a size-k neighborhood using Euclidean distance measure, $|\mathbf{X}_s - \mathbf{X}_j|$ for $j = 1, \cdots, n$ and $j \neq s$;

(2) Obtain the principal eigenvector, $\hat{\zeta}_{\mathbf{X}_s|y}$ and its eigenvalue ω_s in the neighborhood;

(3) Repeat step 1-2 for $s = 1, \cdots, n$;

Estimation Procedure (continue)

(4) Keep m neighborhoods, where $m = 100\delta\% \times n$, and $\delta \in (0, 1)$ representing the fraction of the descending-ordered ω_s for $s = 1, \cdots, n$;

(5) For those neighborhoods that we keep, find the directions in the LCSs.

(6) Combine the local estimates to estimate the CS. The first d eigenvectors corresponding to the first d largest eigenvalues of $\mathbf{V} = \sum_{s=1}^{m} \hat{\beta}_{ys}\hat{\beta}_{ys}^{T}$ form a basis of the CS.

Some issues in the estimation

- The choice of k
  - "Trade off" between a "good direction estimation" and a "good data reduction" (Härdle, 1990);
  - Empirically, $k = p$.
- The selection of neighborhoods
  - Threshold approach;
  - Empirically, $\delta = 0.5$.
- SIR in the neighborhood
  - Partial inverse regression (Li, Cook and Tsai, 2007)

Estimating d

- Permutation test (d=0 and d>0);
- Sphericity test (d>0);
- Bayesian Information Criterion (BIC).

Outline

1. Introduction
2. Localized Sufficient Dimension Reduction
3. KNN Sliced Inverse Regression
4. **Simulation Studies**
5. Summary and Future Directions

Example

Consider the regression model

$$Y = \cos(X_1) + X_2^2 + 0.2\varepsilon$$

In this case we have $\mathcal{S}_{Y|\mathbf{X}} = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$. *Both normal and non-normal predictors are considered. For the upper panel of Table 1* $\mathbf{X}$ *is generated from* $N(\mathbf{0}, \mathcal{I}_p)$. *The the lower panel* $\mathbf{X}$ *is generated by the following scheme*

$$X_1 \sim t_5,\ X_2 \sim F_{3,18},\ X_3 \sim \chi^2_{10},\ X_4 \sim F_{2,12},\ x_5 \sim U(-1,1),\ X_6, \cdots, x_{10} \sim t_8.$$

Table 1: Accuracy of KNN-SIR

Normal predictor							
n	k	p	r	$\boldsymbol{\Delta}(\mathbf{B}, \hat{\mathbf{B}})$	TPR	FPR	% of correct d by Sd
300	10	10	0.916(0.258)	0.511(0.172)	0.967	0.027	59.2
400	10	10	0.966(0.164)	0.424(0.147)	0.987	0.024	71.6
500	10	10	0.993(0.050)	0.363(0.108)	1.000	0.017	77.6
Non-normal predictor							
300	10	10	0.741(0.318)	0.646(0.219)	0.885	0.045	42.4
400	10	10	0.873(0.207)	0.553(0.188)	0.948	0.042	48.8
500	10	10	0.942(0.181)	0.490(0.168)	0.988	0.032	50.2

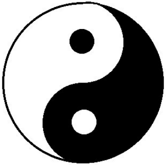
Example

Model: $Y = I[\beta_1^T \mathbf{X} + 0.2\epsilon > 1] + 2I[\beta_2^T \mathbf{X} + 0.2\epsilon > 0]$, *where* $\beta_1 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3 + \mathbf{e}_4$, $\beta_2 = \mathbf{e}_7 + \mathbf{e}_8 + \mathbf{e}_9 + \mathbf{e}_{10}$ *and* $S_{Y|\mathbf{X}} = span(\beta_1, \beta_2)$.

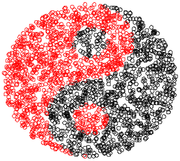
Table 2: Accuracy of KNN-SIR

n	r	$\Delta(\mathbf{B}, \hat{\mathbf{B}})$	% of correct d by Sd
300	0.829(0.157)	0.492(0.164)	58.0
400	0.903(0.068)	0.389(0.116)	69.0
500	0.930(0.049)	0.328(0.109)	80.0

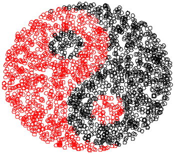
Example: Tai Chi



(a) Tai Chi Figure



(b) Training data in first two predictor directions



(c) Training data projected onto first two KNN-SIR directions

Outline

1. Introduction
2. Localized Sufficient Dimension Reduction
3. KNN Sliced Inverse Regression
4. Simulation Studies
5. Summary and Future Directions

Summary:

1. A novel theory of sufficient dimension reduction based on localization;
2. KNN-SIR and its sparse version;
3. Permutation, sphericity test and BIC for structural dimension d.

Future Directions

1. **KNN-logistic regression and KNN-SVM;**
2. **A unified sphericity test for kernel dimension reduction matrix;**
3. **Applications.**

Thank you!