

# Sliced Inverse Regression For Dimension Reduction (Ker-Chau Li)

Xuelong Wang

January 11, 2018

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression
- 4 Sliced Inverse Regression Method
- 5 Simulation

# On the Agenda

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression
- 4 Sliced Inverse Regression Method
- 5 Simulation

# Regression Analysis

- Study the relationship of a response  $y$  and its covariates  $x$
- Use the information of  $x$  to explain  $y$
- Parametric model
  - Linear Regression model
- Nonparametric model
  - Local smoothing (kNN)

# Curse of Dimensionality

- When the dimension of  $x$  gets higher, observations are far away from each other
- Standard methods probably will break down due to the sparseness of data
- One solution is reducing the dimension of  $x$

# On the Agenda

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression
- 4 Sliced Inverse Regression Method
- 5 Simulation

# A Model for dimension reduction

## Model Settings

$$y = f(\beta_1 x, \dots, \beta_k x, \epsilon)$$

$x$  is explanatory variable, column vectors on  $\mathcal{R}^p$ ,

$\beta'$ s are unknown row vectors,

$\epsilon$  is independent of  $x$ ,

$f$  is an arbitrary unknown function on  $\mathcal{R}^{k+1}$

- $(\beta_1 x, \dots, \beta_k x)'$  is the projection of the  $x \in \mathcal{R}^p$  into  $\mathcal{R}^K$ ,  $K \ll p$
- Captures all we need to know about  $y$

# A Model for dimension reduction

## Effective dimension-reduction

- ① Effective dimension-reduction direction (e.d.r)
    - A Linear combination of  $\beta'$ 's
  - ② A Linear space  $\mathcal{B}$ :
    - Spanned by  $\beta'$ 's ( $\text{Span}(\beta)$ )  $\Leftrightarrow$  All the possible linear combination of  $\beta'$ 's
- 
- Since  $f$  is arbitrary, only the  $\mathcal{B}$  can be identified
  - Inverse Regression is one of the methods of estimating the Effective dimension-reduction space ( $\mathcal{B}$ )



# On the Agenda

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression**
- 4 Sliced Inverse Regression Method
- 5 Simulation

# Inverse Regression

## Inverse Regression

- Regress  $x$  against of  $y$
- Use the information of  $y$  to explain  $x$
- From one  $p$ -dimension problem to  $p$  One-dimension regression problems

# Inverse Regression Curve

## Inverse Regression Curve

$$E(x|y) \in \mathcal{R}^p$$

## Centered Inverse Regression Curve

$$E(x|y) - E(x)$$

- $E[E(x|y)] = E(x)$  is the center
- With certain conditions, the centered inverse curve is related with the e.d.r.!

# Conditions

## Condition 1.1

Conditional Independence

$$y = f(\beta_1 x, \dots, \beta_k x, \epsilon) \Leftrightarrow (y \perp\!\!\!\perp x) | \beta x$$

## Condition 3.1

Linear Condition

For any  $b$  in  $\mathcal{R}^p$ ,

$$E(b\mathbf{x} | \beta_1 \mathbf{x} = \beta_1 x, \dots, \beta_k \mathbf{x} = \beta_k x) = c_0 + c_1 \beta_1 x, \dots, c_k \beta_k x$$

# Centered Inverse Regression Curve and e.d.r

## Theorem 3.1

Under the previous Conditions,

$$E(x|y) - E(x) \subset \text{Span}(\beta_k \Sigma_{xx}), k = 1, \dots, K$$

The centered inverse regression curve is contained in the linear subspace spanned by  $\beta_k \Sigma_{xx}$

# Centered Inverse Regression Curve and e.d.r

## Corollary 3.1

$$z = \Sigma_{xx}^{-1/2}[x - E(x)]$$

$x$  is the standardized

$$f(\beta_1 x, \dots, \beta_k x, \epsilon) \Rightarrow f(\eta_1 z, \dots, \eta_k z, \epsilon) \Rightarrow \beta_k = \eta_k \Sigma_{xx}^{-1/2}$$

$$E(z|y) - E(z) \subset \text{Span}(\eta_k), k = 1, \dots, K$$

# An Important consequence

## Covariance matrix is the key

- The Covariance matrix  $\text{Cov}(E(z|y))$  is degenerated in any direction which is orthogonal to  $\eta'$ s
- $\eta_k$ 's ( $k = 1, \dots, K$ ) associated with largest  $K$  eigenvalues of  $\text{Cov}(E(z|y))$

## How to estimate the $\text{Cov}(E(z|y))$

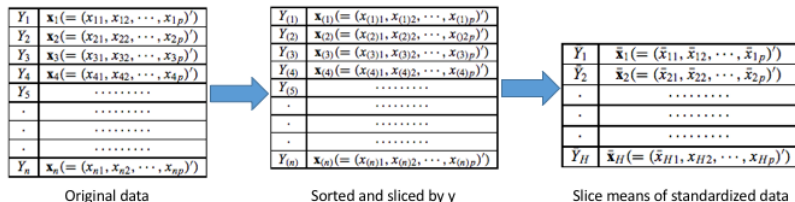
That leads to Sliced Inverse Regression Method

# On the Agenda

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression
- 4 Sliced Inverse Regression Method**
- 5 Simulation



# Slice Inverse Regression Method



$$\hat{V} = n^{-1} \sum_{h=1}^H n_h \bar{x}_h \bar{x}_h^T \quad \longrightarrow \quad \begin{array}{l} \text{Conduct PCA on } \hat{V} \\ \text{Find the first Kth eigenvectors } \hat{\eta} \end{array} \quad \longrightarrow \quad \hat{\beta}_k = \hat{\eta}_k \Sigma_{xx}^{-1/2}$$

Estimated Covariance matrix

# Sliced Inverse Regression Method with details

- 1 Standardize  $x$ 
  - $z_i = \Sigma_{xx}^{-1/2}(x_i - \bar{x})(i = 1, \dots, n)$
- 2 Divide the range of  $y$  into  $H$  slices,  $I_1, \dots, I_H$ 
  - $\hat{p}_h = (1/n) \sum_{i \in I_h} 1$
- 3 Calculate the sample mean for each slice
  - $\hat{m}_h = (1/n\hat{p}_h) \sum_{i \in I_h} z_i$
- 4 Conduct a Principal Component Analysis on the estimated Covariance matrix
  - $\hat{V} = \sum_{h=1}^H \hat{p}_h \hat{m}_h \hat{m}_h'$
- 5 Select the  $K$  largest eigenvectors (row vectors)
  - $\hat{\eta}_k (k = 1, \dots, K)$
- 6 Transform the eigenvectors back to original scale
  - $\hat{\beta}_k = \hat{\eta}_k \hat{\Sigma}_{xx}^{-1/2}$

# On the Agenda

- 1 Introduction
- 2 A Model for dimension reduction
- 3 Inverse Regression
- 4 Sliced Inverse Regression Method
- 5 Simulation

# Simulation 1

## Simulation settings

$$y = x_1 + x_2 + x_3 + x_4 + 0x_5 + \epsilon$$

- $n = 100, p = 5$
- Only one component  $\beta = (1, 1, 1, 1, 0)$
- Normalized target  $\beta^* = (0.5, 0.5, 0.5, 0.5, 0)$
- The number of slice  $H = (5, 10, 20)$

# Simulation 1 results

*Table 1. Mean and Standard Deviation\* of  $\hat{\beta}_1 = (\hat{\beta}_{11}, \dots, \hat{\beta}_{15})$  for the linear model (6.1),  $n = 100$ ; the Target is (.5, .5, .5, .5, 0)*

$H$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{14}$	$\hat{\beta}_{15}$
5	.505 (.052)	.498 (.049)	.494 (.056)	.488 (.056)	.002 (.066)
10	.502 (.046)	.500 (.045)	.492 (.055)	.491 (.049)	.001 (.060)
20	.500 (.048)	.502 (.046)	.497 (.053)	.487 (.054)	-.003 (.060)

\*Numbers in parentheses represent standard deviations.

- Repeat the simulation 100 times to generate the empirical distribution of  $\hat{\beta}'$ s
- $\beta$  and  $\hat{\beta}$  are standardized

# Simulation 2

## Simulation settings

$$y = x_1(x_1 + x_2 + 1) + \sigma \cdot \epsilon$$

$$y = \frac{x_1}{0.5 + (x_2 + 1.5)^2} + \sigma \cdot \epsilon$$

- $n = 400$ ,  $p = 10$
- $\sigma = (0.5, 1)$
- The number of slice  $H = (5, 10, 20)$
- Two components  $\beta_1 = (1, 0, 0, \dots, 0)$ ,  $\beta_2 = (0, 1, 0, \dots, 0)$

# Evaluate the effectiveness of estimated e.d.r direction

## Criterion of one direction

$$R^2(\hat{b}) = \max_{\beta \in \mathcal{B}} \frac{(\hat{b} \Sigma_{xx} \beta')^2}{\hat{b} \Sigma_{xx} \hat{b}' \cdot \beta \Sigma_{xx} \beta'}$$

Squared correlation coefficient between the  $\hat{b}x$  and  $\beta_1 x, \dots, \beta_k x$   
Invariant under affine transformation of  $x$

## Criterion of the subspace $\mathcal{B}$

$$R^2(\hat{\mathcal{B}}) = \frac{\sum_{k=1}^K R^2(\hat{b}_k)}{K}$$

## Simulation 2 results

*Table 2. Mean and Standard Deviation of  $R^2(\hat{\beta}_1)$  and  $R^2(\hat{\beta}_2)$  for the Quadratic Model (6.2),  $p = 10$ ,  $n = 400$*

$H$	$\sigma = 0.5$		$\sigma = 1$	
	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$
5	.91 (.05)	.75 (.15)	.88 (.07)	.52 (.21)
10	.92 (.04)	.80 (.13)	.89 (.08)	.55 (.24)
20	.93 (.04)	.77 (.15)	.88 (.08)	.49 (.26)

- Repeat the simulation 100 times to generate the empirical distribution of  $\beta'$ s



## Simulation 3 results

*Table 3. Mean and Standard Deviation of  $R^2(\hat{\beta}_1)$  and  $R^2(\hat{\beta}_2)$  for the Rational Function Model (6.3),  $p = 10$ ,  $n = 400$*

$H$	$\sigma = 0.5$		$\sigma = 1$	
	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$
5	.96	.83	.89	.51
	(.02)	(.08)	(.06)	(.23)
10	.96	.88	.90	.56
	(.02)	(.06)	(.06)	(.23)
20	.96	.89	.90	.53
	(.02)	(.06)	(.06)	(.24)

- Repeat the simulation 100 times to generate the empirical distribution of  $\beta'$ s

# Graphics

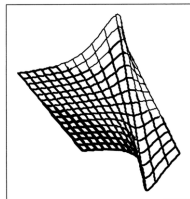


Figure 2: Response surface of simulation 3

How to visualize the data?

Directions found by SIR

# Thank you

# Reference

Li, Ker-Chau. 1991. "Sliced Inverse Regression for Dimension Reduction."