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

Xuelong Wang

January 10, 2018

Regression Analysis

- Study the relationship of a response variable y and its explanatory variable 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
- We need to reduce the dimension of x so that it's easier for visualizing data and fitting models

A Model for dimension reduction

Model Settings

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

x is explanatory variable, column vectors on $\mathcal{R}^p$,
$\beta' s$ are unknown row vectors,
ϵ is independent of x,
f is an arbitrary unknown function on $\mathcal{R}^{p+1}$

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

A Model for dimension reduction

Effective dimension-reduction

1. Effective dimension-reduction directions (e.d.r)
   - Any Linear combination of $\beta' s$
2. A Linear space $\mathcal{B}$:
   - Spanned by $\beta' s \Leftrightarrow \ \ Span(\beta)$

- Since f is arbitrary, only the $\mathcal{B}$ can be identified
- Inverse Regression one of the methods of estimating the Effective dimension-reduction directions

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

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

Condition 3.1

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

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

Centered Inverse Regression Curve and e.d.r

Theorem 3.1

Under the previous Conditions,

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

The centered inverse regression curve is contained in the linear subsapce 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

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

$$E(z|y) - E(z) \subset Span(\eta_k), k = 1, \ldots, K$$

An Important consequence

Covariance matrix is the key

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

How to estimate the $Cov(E(z|y))$

That leads to Sliced Inverse Regression Method

Sliced Inverse Regression Method

1. Standardize x
   - $z_i = \Sigma_{xx}^{-1/2}(x_i - \bar{x})(i = 1, \ldots, n)$
2. Divide the range of y into H slices, $I_1, \ldots, I_H$
   - $\hat{p}_h = (1/n)\sum_{i=1}^{n}(I_{y_i \in I_h})$
3. Calculate the sample mean for each slice
   - $\hat{m}_h = (1/n\hat{p}_h)\sum_{y_i \in I_h} z_i$
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)
   - $\hat{\eta}_k (k = 1, ..., K)$
6. Transform the eigenvectors back to original scale
   - $\hat{\beta}_k = \hat{\eta}_k \hat{\Sigma}_{xx}^{-1/2}$

Simulation 1

Simulation settings

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

- $n = 100$
- 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 $\hat{\beta}_1 = (\hat{\beta}_{11}, \ldots, \hat{\beta}_{15})$ 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	.498	.494	.488	.002
	(.052)	(.049)	(.056)	(.056)	(.066)
10	.502	.500	.492	.491	.001
	(.046)	(.045)	(.055)	(.049)	(.060)
20	.500	.502	.497	.487	−.003
	(.048)	(.046)	(.053)	(.054)	(.060)

*Numbers in parentheses represent standard deviations.

- Repeat the simulation 100 times to generate the empirical distribution of $\hat{\beta}'s$
- Since we want to compare two directions, β and $\hat{\beta}$ are standardized

Simulation 2

Simulation settings

$$y = x_1(x_1 + x_2 + 1) + \sigma \cdot \epsilon$$
$$y = \frac{x_1}{0.5 + (x_2 + 1.5)^2} + \sigma \cdot \epsilon$$

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

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 bx and $\beta_1 x, \ldots, \beta_k x$

Invariant under affine transformation of x

Criterion of the subspace 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), $p = 10$, $n = 400$

	$\sigma = 0.5$		$\sigma = 1$	
H	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$
5	.91	.75	.88	.52
	(.05)	(.15)	(.07)	(.21)
10	.92	.80	.89	.55
	(.04)	(.13)	(.08)	(.24)
20	.93	.77	.88	.49
	(.04)	(.15)	(.08)	(.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), $p = 10$, $n = 400$

	$\sigma = 0.5$		$\sigma = 1$	
H	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$	$R^2(\hat{\beta}_1)$	$R^2(\hat{\beta}_2)$
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 $\beta' s$

Thank you

Reference

Tonglin Zhang, Baijian Yang. 2016. “Big Data Dimension Reduction Using Pca.”