ST720 Data Science

Sufficient Dimension Reduction

Seung Jun Shin (sjshin@krea.ac.kr)

Department of Statistics, Korea University

Dimension Reduction in Regression

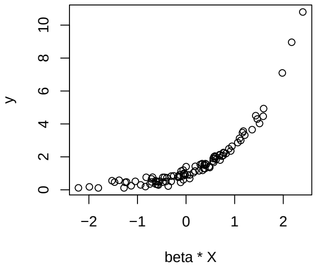
- Idea: First, reduce dimension of $\mathbf{X}$, then build a regression model one the reduced space.
- ex) Principal component (PC) regression
  1. Apply PCA to $\mathbf{X}$. (Dimension Reduction)
  2. Regression of Y on the first few PCs. (Regression)

Motivating Example 1: PC Regression

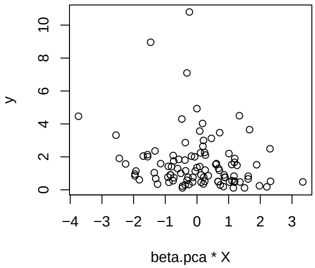
- Simple model:

$$y_i = \exp(\beta^T \mathbf{x}_i) + \epsilon_i, \qquad i = 1, \cdots, 100$$

where $p = 5, \beta = (1, 0, 0, 0, 0)$, and $x_{ij} \sim N(0, 1), j = 1, \cdots, p$.



(a) True



(b) PCA

Motivating Example 1: Sliced Inverse Regression

- We applied the sliced inverse regression (SIR) to the previous simple example.

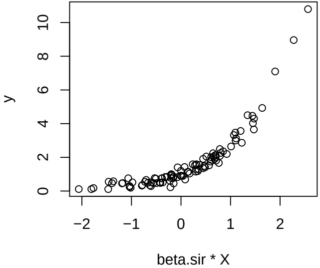


Figure 1: Optimal λ selection

	β	$\hat{\beta}_{PCA}$	$\hat{\beta}_{SIR}$
X_1	1	-0.23	1.00
X_2	0	-0.26	-0.02
X_3	0	-0.07	-0.04
X_4	0	-0.02	-0.08
X_5	0	-0.94	-0.03

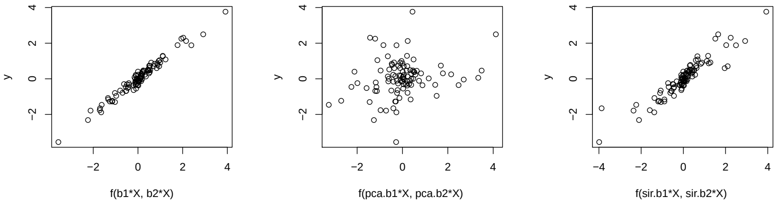
Table 1: Estimated β

Motivating Example 2

- Consider

$$y_i = \frac{\beta_1^T \mathbf{x}_i}{0.5 + (\beta_2^T \mathbf{x}_i + 1)^2} + \epsilon_i, \qquad i = 1, \cdots, 100$$

where $\beta_1 = (1, 0, 0, 0, 0)$ and $\beta_2 = (0, 1, 0, 0, 0)$.



(a) True (b) PCA (c) SIR

Figure 2: Scatter Plots of Y vs $f(\beta_1^T \mathbf{X}, \beta_2^T \mathbf{X}) = X_1/\{0.5 + (X_2 + 1)^2\}$

.

Motivating Example 2

- Results

	β_1	$\hat{\beta}_{1,PCA}$	$\hat{\beta}_{1,SIR}$	β_2	$\hat{\beta}_{2,PCA}$	$\hat{\beta}_{2,SIR}$
X_1	1	-0.23	1.00	0	0.03	0.06
X_2	0	-0.26	-0.01	1	0.27	-0.95
X_3	0	-0.07	-0.03	0	-0.78	0.01
X_4	0	-0.02	0.05	0	-0.57	-0.14
X_5	0	-0.94	0.04	0	-0.01	-0.26

Table 2: Estimated β

Sufficient Dimension Reduction

For a given pair of $(Y, \mathbf{X}) \in \mathbb{R} \times \mathbb{R}^p$,

- Sufficient Dimension Reduction (SDR) seeks a matrix $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_d) \in \mathbb{R}^{p \times d}$ which satisfies

$$Y \perp \mathbf{X} | \mathbf{B}^\top \mathbf{X}, \tag{1}$$

  which is equivalent to

  - $Y|\mathbf{X} \stackrel{\mathcal{D}}{=} Y|\mathbf{B}^\top \mathbf{X}$.
    : If $d \ll p$, dimension reduction is achieved.
  - $\mathbf{X}|(Y, \mathbf{B}^\top \mathbf{X}) \stackrel{\mathcal{D}}{=} \mathbf{X}|\mathbf{B}^\top \mathbf{X}$.
    : $T(\mathbf{X})$ is a sufficient statistic for a parameter θ *if and only if*

$$\mathbf{X}|(\theta, T(\mathbf{X})) \stackrel{\mathcal{D}}{=} \mathbf{X}|T(\mathbf{X}).$$

Central Subspace: Identifiable Target of Interest in SDR

Dimension Reduction Subspace (DRS):

- Any subspace $\mathrm{span}(\mathbf{B}) \subseteq \mathbb{R}^p$ that satisfies (1).
- Not unique as $\mathbf{B}$.
- Require uniqueness/minimality.

Central Subspace

Central Subspace, $\mathcal{S}_{Y|\mathbf{X}}$ is the intersection of all DRSes.

- $\mathcal{S}_{Y|\mathbf{X}}$ has minimum dimension among all DRS and uniquely exists under very mild conditions.
- We assume $\mathcal{S}_{Y|\mathbf{X}} = \mathrm{span}(\mathbf{B})$ and $\mathbf{B}$ is unique in this sense.
- The dimension of $\mathcal{S}_{Y|\mathbf{X}}$, d is called the structural dimension.

Estimation of $\mathbf{B}$ or $\mathcal{S}_{Y|\mathbf{x}}$

- K-C Li (1991) Sliced Inverse Regression for Dimension Reduction.
- A lot of methods follows up to today,
  - Sliced Average Variance Estimation (SAVE, 1991)
  - Principal Hessian Directions (pHd, 1992)
  - Directional Regression (2007)
  - Principle Support Vector Machine (2011)
  - ...
- `dr` package avaiable in `R`.

Sliced Inverse Regression

- Suppose $\mathbf{X}$ is standardized with $E(\mathbf{X}) = \mathbf{0}$ and $\mathrm{var}(\mathbf{X}) = \Sigma$.
- Then we have (under a certain condition)

$$\mathbb{E}(\mathbf{X}|Y) \in \mathcal{S}_{Y|\mathbf{X}}.$$

- Let's apply PCA to $\mathbb{E}(\mathbf{X}|Y)$ which can be estimated by sliced means:

$$\hat{\mathbb{E}}(\mathbf{X}|Y) = n_h^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbb{1}\{i \in I_h\}$$

where $I_h = \{i : l_h < y_i < u_h\}$ and $n_h = |I_h|, h = 1, \cdots, H$

Example: SIR.R

```
library(dr)
data(ais)

y <- ais$LBM
x <- cbind(ais$SSF, ais$Wt,  ais$Hg, ais$Ht,
           ais$WCC, ais$RCC, ais$Hc, ais$Ferr)
x <- log(x) # take log
obj.sir <- dr(y ~ x, method = "sir")
```

Example: SIR.R

```r
obj.sir$evectors[,1:2]
```

```
##             Dir1          Dir2
## x1  0.150963358 -0.0501785457
## x2 -0.916480522 -0.1942298625
## x3 -0.131538894  0.6854750758
## x4 -0.093358860 -0.0433408964
## x5  0.004467838  0.0001833808
## x6 -0.188973540  0.3475652934
## x7  0.274758965 -0.6058301419
## x8 -0.005631238  0.0130588502
```

```r
summary(obj.sir)$test
```

```
##                     Stat df      p.value
## 0D vs >= 1D 298.91506 80 0.0000000000
## 1D vs >= 2D 105.46741 63 0.0006413834
## 2D vs >= 3D  55.96900 48 0.2006081867
## 3D vs >= 4D  34.33966 35 0.4998059566
```

Example: SIR.R

- Structural Dimension

```
summary(obj.sir)$test
```

```
##                         Stat df      p.value
## 0D vs >= 1D 298.91506 80 0.0000000000
## 1D vs >= 2D 105.46741 63 0.0006413834
## 2D vs >= 3D  55.96900 48 0.2006081867
## 3D vs >= 4D  34.33966 35 0.4998059566
```

Example: SIR.R

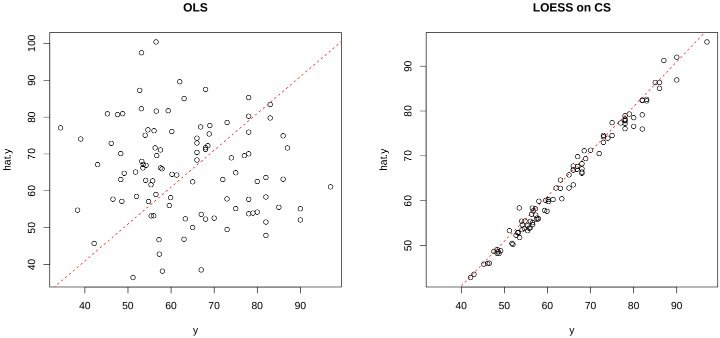


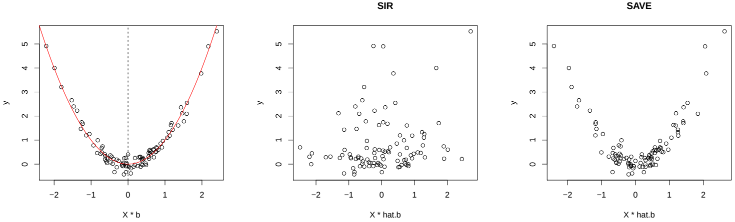
Figure 3: test response vs estimated one: (a) Linear regression on the original space. (b) Nonparametric regression (LOESS) on the estimated central subspace. Note that LOESS on the original space is not applicable due to the dimensionality of X.

SAVE: Motivating Example

- Consider

$$y_i = (\beta^T \mathbf{x}_i)^2 + \epsilon_i \qquad i = 1, \cdots, n,$$

where $n = 100$, $p = 5$ and $\beta = (1, 0, 0, 0, 0)$.



(a) True (b) SIR (c) SAVE

Figure 4: Scatter Plots of Y vs $\beta^T \mathbf{X}$

.

Why SIR fails?

- SIR fails when the regression function is symmetric about the origin, since $E(\mathbf{X}|Y) = 0$ for all Y.

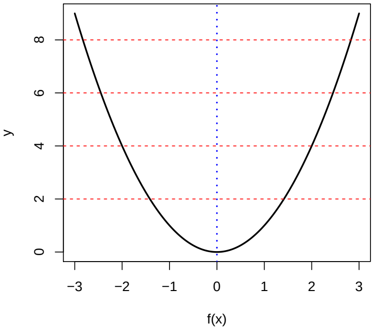


Figure 5: Symmetric regression function

Example: SAVE.R

```
obj.save <- dr(y ~ x, method = "save")
obj.save
```

```
##
##  dr(formula = y ~ x, method = "save")
## Estimated Basis Vectors for Central Subspace:
##            Dir1        Dir2         Dir3        Dir4
## x1  0.136221335 -0.01326124 -0.003848012  0.03937275
## x2 -0.933124616 -0.09136606  0.145207508 -0.29330884
## x3 -0.193297576  0.50057993  0.300768199  0.30829737
## x4  0.195232876  0.39625072 -0.614591360  0.47226293
## x5  0.019005007  0.01961664 -0.015912595 -0.06058359
## x6 -0.087529966  0.14530993  0.338924072  0.56061509
## x7  0.164905792 -0.74985379 -0.628710946 -0.52096816
## x8  0.005075835 -0.01004954 -0.017768863  0.07051725
## Eigenvalues:
## [1] 0.9477429 0.6432531 0.5862137 0.4366803
```

Example: SAVE.R

- Structural Dimension

```
dr.permutation.test(obj.save, npermute = 100)
```

```
## $summary
##                       Stat    p.value
## 0D vs >= 1D 803.9896 0.08910891
## 1D vs >= 2D 612.5456 0.19801980
## 2D vs >= 3D 482.6084 0.10891089
## 3D vs >= 4D 364.1933 0.29702970
## 4D vs >= 5D 275.9839 0.32673267
##
## $npermute
## [1] 100
##
## attr(,"class")
## [1] "dr.permutation.test"
```

Practical Consideration

- Number of slices
- SIR or SAVe?
- Determination of structural dimension
- Subsequent regression function estimator.