On multiple-case diagnostics in linear subspace method

Kuniyoshi HAYASHI[1] Hiroyuki MINAMI[2] Masahiro MIZUTA[2]

[1]Graduate School of Information Science and Technology, Hokkaido University
[2]Information Initiative Center, Hokkaido University

Outline

1. Introduction
2. Sensitivity analysis in linear subspace method
3. A multiple-case diagnostics with clustering
4. Numerical example
5. Concluding remarks

Outline

Introduction

Sensitivity analysis in pattern recognition

Single-case diagnostics in linear subspace method (Hayashi *et al.*, 2008)

Multiple-case diagnostics in linear subspace method

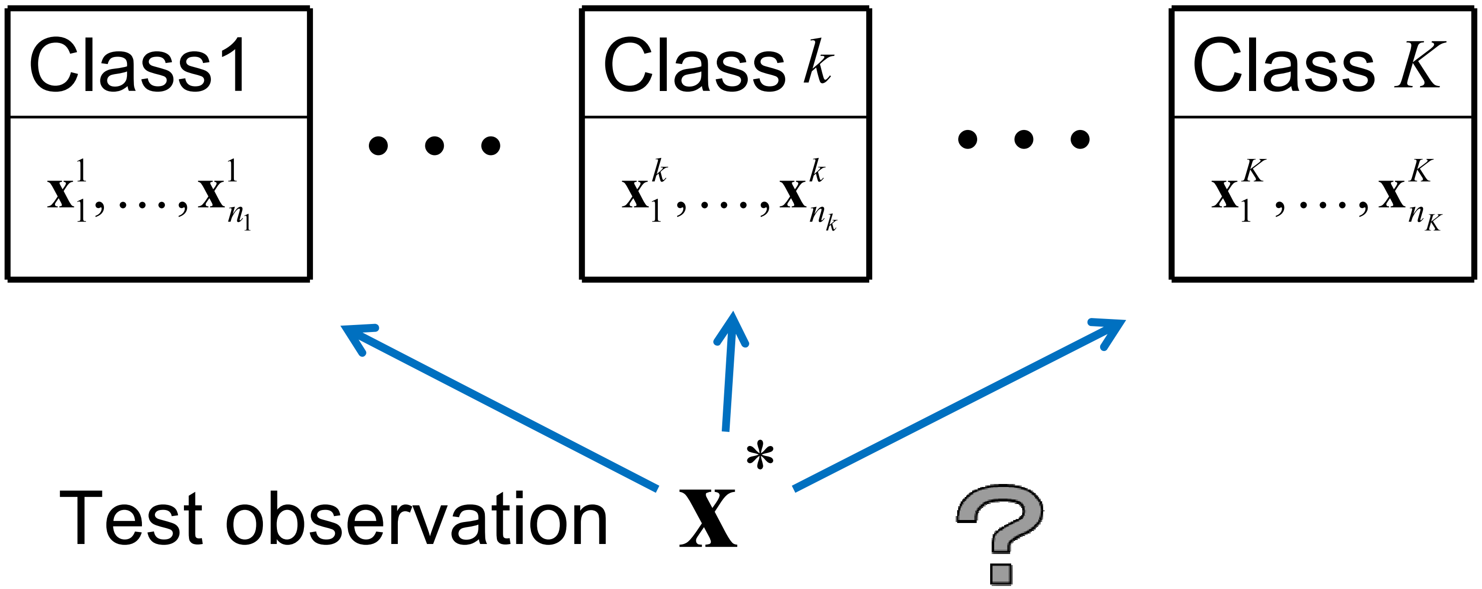
We show the availability through a simulation study.

Outline

Linear subspace method

CLAss-Featuring Information Compression (CLAFIC; Watanabe, 1967)

Training observations $\mathbf{x}_i^k \in R^p \ (i = 1, \ldots, n_k ; k = 1, \ldots, K)$



[CLAFIC]

Autocorrelation matrix of the training data in k-th class :

$$\hat{G}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^k \mathbf{x}_i^{k\,\mathrm{T}}$$

Eigenvalues : $\hat{\lambda}_1^k \geq \hat{\lambda}_2^k \geq \cdots \geq \hat{\lambda}_p^k \geq 0$

Eigenvectors : $\hat{\mathbf{u}}_s^k \ (s = 1, \ldots, p)$

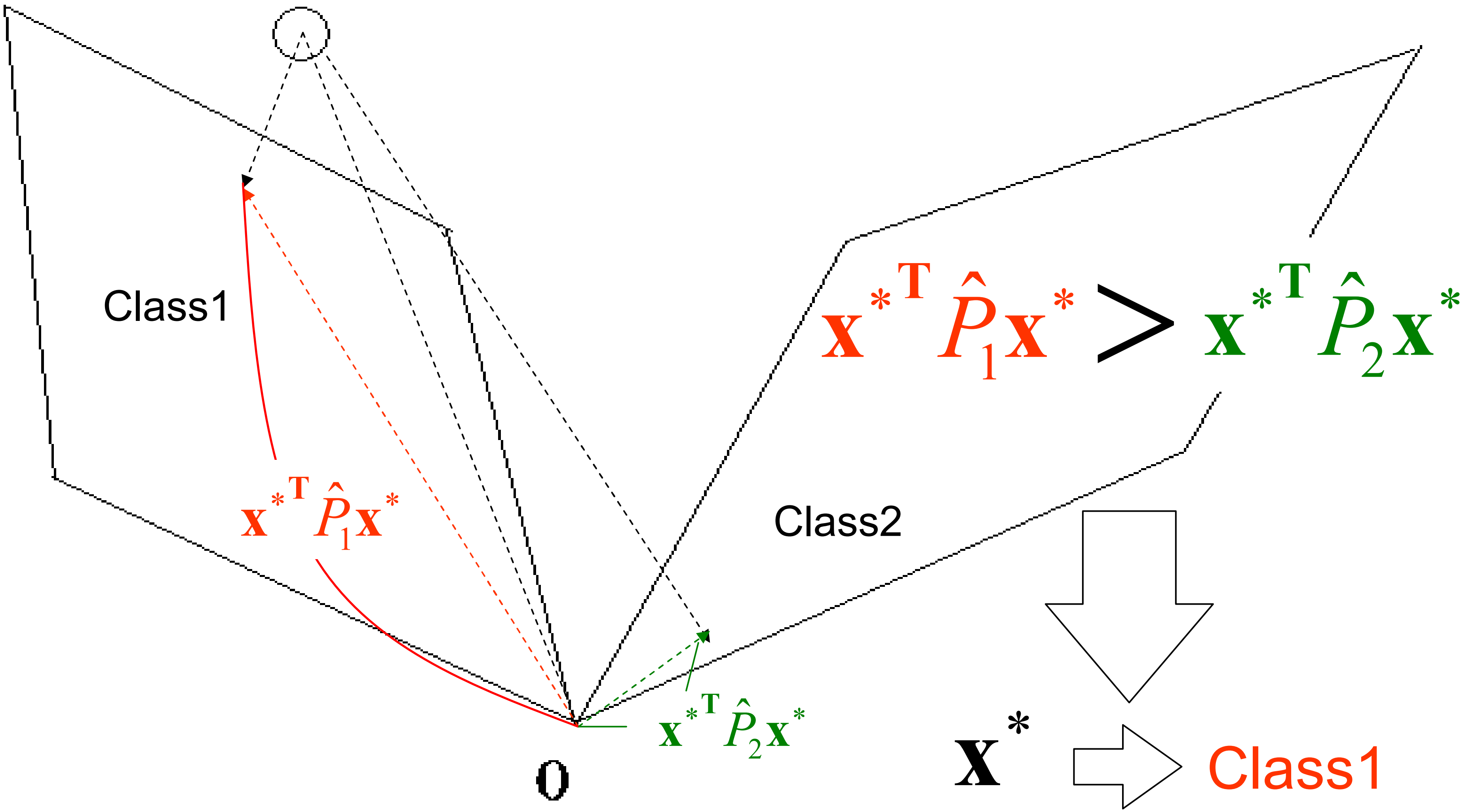
A projection matrix in k-th class:

$$\hat{P}_k = \sum_{s=1}^{p_k} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\,\mathrm{T}} \quad (1 \leq p_k \leq p)$$

$$k := \arg\max \left\{ \mathbf{x}^{*\mathrm{T}} \hat{P}_k \mathbf{x}^* \right\}$$

CLAFIC

$\mathbf{x}^*$ Test observation



[Discriminant score and Average]

Discriminant score for $\mathbf{x}_i^k$:

$$\hat{z}_i^k = \mathbf{x}_i^{k\,\mathbf{T}} \hat{Q}_k \mathbf{x}_i^k \qquad (1 \le i \le n_k),$$

where $$\hat{Q}_k = \frac{1}{K-1}\left(K\hat{P}_k - \sum_{l=1}^{K} \hat{P}_l \right).$$

Average discriminant score

$$\hat{Z}^k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{z}_i^k \ .$$

[Sample influence function (SIF)]

$$\mathbf{SIF}\left(\mathbf{x}_j^g;\mathbf{vech}\left(\hat{Q}_k\right)\right)=-\left(n_g-1\right)\cdot\left\{\mathbf{vech}\left(\hat{Q}_{k\,g(j)}\right)-\mathbf{vech}\left(\hat{Q}_k\right)\right\},$$

$\mathbf{vech}\left(\hat{Q}_{k\,g(j)}\right)$ means $\mathbf{vech}\left(\hat{Q}_k\right)$ without j-th observation in g-th class

[Empirical influence function (EIF)]

$$\mathbf{EIF}\left(\mathbf{x}_j^g;\mathbf{vech}\left(\hat{Q}_k\right)\right)=\begin{cases}\mathbf{vech}\left(\displaystyle\sum_{s=1}^{p_g}\sum_{t=p_g+1}^{p}\left(\hat{\lambda}_s^g-\hat{\lambda}_t^g\right)^{-1}\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\hat{G}_g^{g\,j}\hat{\mathbf{u}}_t^g\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_t^{g\,\mathbf{T}}+\hat{\mathbf{u}}_t^g\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\right)\right) & (g=k)\\ \mathbf{vech}\left(-\dfrac{1}{K-1}\displaystyle\sum_{s=1}^{p_g}\sum_{t=p_g+1}^{p}\left(\hat{\lambda}_s^g-\hat{\lambda}_t^g\right)^{-1}\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\hat{G}_g^{g\,j}\hat{\mathbf{u}}_t^g\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_t^{g\,\mathbf{T}}+\hat{\mathbf{u}}_t^g\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\right)\right) & (g\neq k)\end{cases},$$

where p_g is the number of the basis vectors in g-th class.

Case of $\mathbf{SIF}\left(\mathbf{x}_j^g;\mathbf{vech}\left(\hat{Q}_k\right)\right)$

$$\hat{Z}_k^{g(j)} = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbf{x}_i^{k\,\mathbf{T}} \hat{Q}_k^{g(j)} \mathbf{x}_i^k ,$$

$$\hat{Q}_k^{g(j)} = -\left(n_g - 1\right)\cdot\left(\hat{Q}_{k\,g(j)} - \hat{Q}_k\right).$$

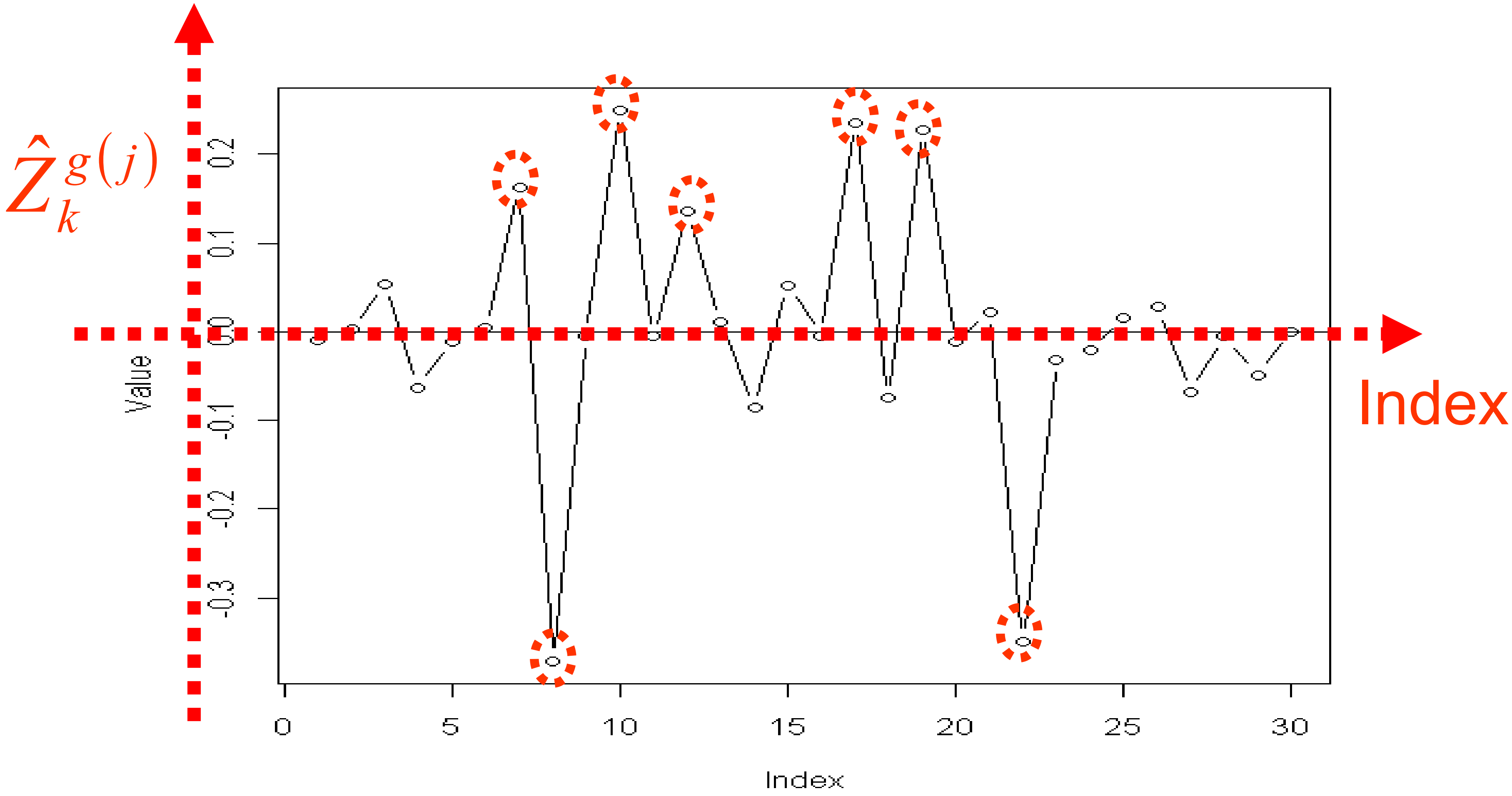
Case of $\mathbf{EIF}\left(\mathbf{x}_j^g;\mathbf{vech}\left(\hat{Q}_k\right)\right)$

$$\hat{Z}_k^{g\,j} = \frac{1}{n_k}\sum_{i=1}^{n_k} \mathbf{x}_i^{k\,\mathbf{T}} \hat{Q}_k^{g\,j} \mathbf{x}_i^k \quad \left(g \neq k\right),$$

$$\hat{Q}_k^{g\,j} = -\frac{1}{K-1}\sum_{s=1}^{p_g}\sum_{t=p_g+1}^{p}\left(\hat{\lambda}_s^g - \hat{\lambda}_t^g\right)^{-1}\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\hat{G}_g^{g\,j}\hat{\mathbf{u}}_t^g\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_t^{g\,\mathbf{T}} + \hat{\mathbf{u}}_t^g\hat{\mathbf{u}}_s^{g\,\mathbf{T}}\right).$$

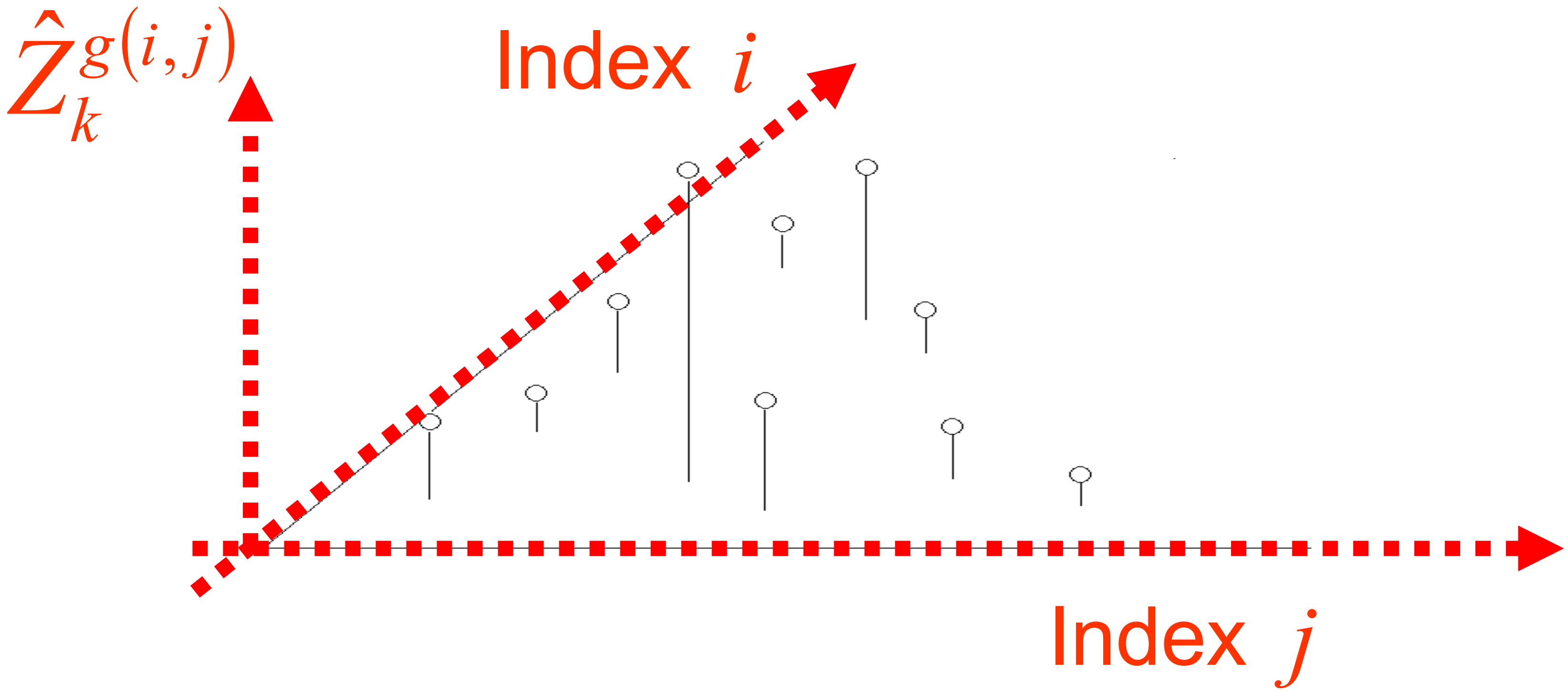
Single-case diagnostics

Influence of single observation for $\hat{Z}^k$



Multiple-case diagnostics

Influence of multiple observations for $\hat{Z}^k$ with $\hat{Z}_k^{g(i,j)} \left(i=1,\ldots,n_g\,;\, j=1,\ldots,n_g\right)$



Multiple-case diagnostics

$\mathbf{w}_0^g = (1,1,\ldots,1)^{\mathrm{T}}$	Weights of unperturbed observations in g-th class
$\mathbf{vech}\left(\hat{Q}_k\right)$	Maximum likelihood estimator
$L\left(\mathbf{vech}\left(\hat{Q}_k\right) \middle\| \mathbf{w}_0^g\right)$	Maximum log likelihood
$\mathbf{w}^g = \mathbf{w}_0^g + t\mathbf{h}$	Weights of perturbed case in g-th class $\lVert\mathbf{h}\rVert = 1$
$L\left(\mathbf{vech}\left(\hat{Q}_{k\mathbf{w}^g}\right) \middle\| \mathbf{w}^g\right)$	Maximum log likelihood in g-th perturbed class

Based on Cook (1986) and Tanaka (1994),

$$D(\mathbf{w}^g) = 2[L(\mathbf{vech}(\hat{Q}_k)|\mathbf{w}_0^g) - L(\mathbf{vech}(\hat{Q}_{k\mathbf{w}^g})|\mathbf{w}_0^g)]$$

$$D(t) = D(0) + \frac{dD(0)}{dt}t + \frac{1}{2}\frac{d^2D(0)}{dt^2}t^2 + \mathrm{O}(t^3).$$

$$D(t) \cong (Const)\cdot \mathbf{h}^{\mathbf{T}}[-\frac{\partial^2 L}{\partial \mathbf{w}^g\ \partial \mathbf{w}^{g\,\mathbf{T}}}\bigg|_{\mathbf{w}^g=\mathbf{w}_0^g}]\mathbf{h}\cdot t^2 .$$

$$C_{\mathbf{h}} = \mathbf{h}^{\mathbf{T}}[-\frac{\partial^2 L}{\partial \mathbf{w}^g\ \partial \mathbf{w}^{g\,\mathbf{T}}}\bigg|_{\mathbf{w}^g=\mathbf{w}_0^g}]\mathbf{h}.$$

$$C_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}}\left[\frac{\partial\,\mathbf{vech}\left(\hat{Q}_{k\,\mathbf{w}^g}\right)^{\mathrm{T}}}{\partial\,\mathbf{w}^g}\right]\left[-\frac{\partial^2 L}{\partial\,\mathbf{vech}(Q_k)\partial\,\mathbf{vech}(Q_k)^{\mathrm{T}}}\right]\left[\frac{\partial\,\mathbf{vech}\left(\hat{Q}_{k\,\mathbf{w}^g}\right)}{\partial\,\mathbf{w}^g}\right]\mathbf{h}.$$

$[\partial\,\mathbf{vech}\left(\hat{Q}_{k\,\mathbf{w}^g}\right)^{\mathrm{T}}/\partial\,\mathbf{w}^g]$ and $[-\partial^2 L/\partial\,\mathbf{vech}\,(Q_k)\partial\,\mathbf{vech}\,(Q_k)^{\mathrm{T}}]$ are evaluated at $\mathbf{w}^g = \mathbf{w}_0^g$ and $\mathbf{vech}(Q_k) = \mathbf{vech}\left(\hat{Q}_k\right)$, respectively.

By $\mathbf{EIF}\left(\mathbf{x}_r^g;\mathbf{vech}\left(\hat{Q}_k\right)\right) = n_g\cdot\left.\frac{\partial\,\mathbf{vech}\left(\hat{Q}_{k\,\mathbf{w}^g}\right)}{\partial\,\mathbf{w}_r^g}\right|_{\mathbf{w}_r^g=0}$ and $\mathbf{acov}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right) = \left\{-E\left[\frac{\partial^2 L}{\partial\,\mathbf{vech}(Q_k)\partial\,\mathbf{vech}(Q_k)^{\mathrm{T}}}\right]\right\}^{-1}$,

we estimate $C_{\mathbf{h}}$ as

$$C_{\mathbf{h}} \cong \mathbf{h}^{\mathrm{T}}[\mathbf{EIF}_{k\,g}][\hat{\mathbf{acov}}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right)]^{-1}[\mathbf{EIF}_{k\,g}]^{\mathrm{T}}\mathbf{h},$$

where $\left[\mathbf{EIF}_{k\,g}\right] = \left\{\mathbf{EIF}\left(\mathbf{x}_r^g;\mathbf{vech}\left(\hat{Q}_k\right)\right)\right\}$ and $\hat{\mathbf{acov}}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right) = \left\{-\frac{\partial^2 L}{\partial\,\mathbf{vech}\left(\hat{Q}_k\right)\partial\,\mathbf{vech}\left(\hat{Q}_k\right)^{\mathrm{T}}}\right\}$.

We solve the following eigenvalue problem.

$$\left[\underbrace{[\mathbf{EIF}_{kg}]\left[\hat{\mathbf{acov}}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right)\right]^{-1}[\mathbf{EIF}_{kg}]^{\mathbf{T}}}_{C_{\mathbf{h}}\text{ : Cook's Curvature}} - \lambda \mathbf{I} \right] \mathbf{h} = 0.$$

Here, we estimate $\hat{\mathbf{acov}}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right)$ with jackknife method.

$$\hat{\mathbf{acov}}\left(\mathbf{vech}\left(\hat{Q}_k\right)\right) \cong V_{\mathbf{JACK}/k}$$

$$= \frac{N-1}{N} \sum_{i=1}^{N} \left[\mathbf{vech}\left(\hat{Q}_k\right)_{N-1,i} - \frac{1}{N} \sum_{j=1}^{N} \mathbf{vech}\left(\hat{Q}_k\right)_{N-1,j} \right] \left[\mathbf{vech}\left(\hat{Q}_k\right)_{N-1,i} - \frac{1}{N} \sum_{j=1}^{N} \mathbf{vech}\left(\hat{Q}_k\right)_{N-1,j} \right]^{\mathbf{T}},$$

where we put $\mathbf{vech}\left(\hat{Q}_k\right)_{N-1,r}$ $(r = 1, \ldots, N)$ as $\hat{Q}_k^{g(r)} - \hat{Q}_k$.

1 Get the eigenvectors with

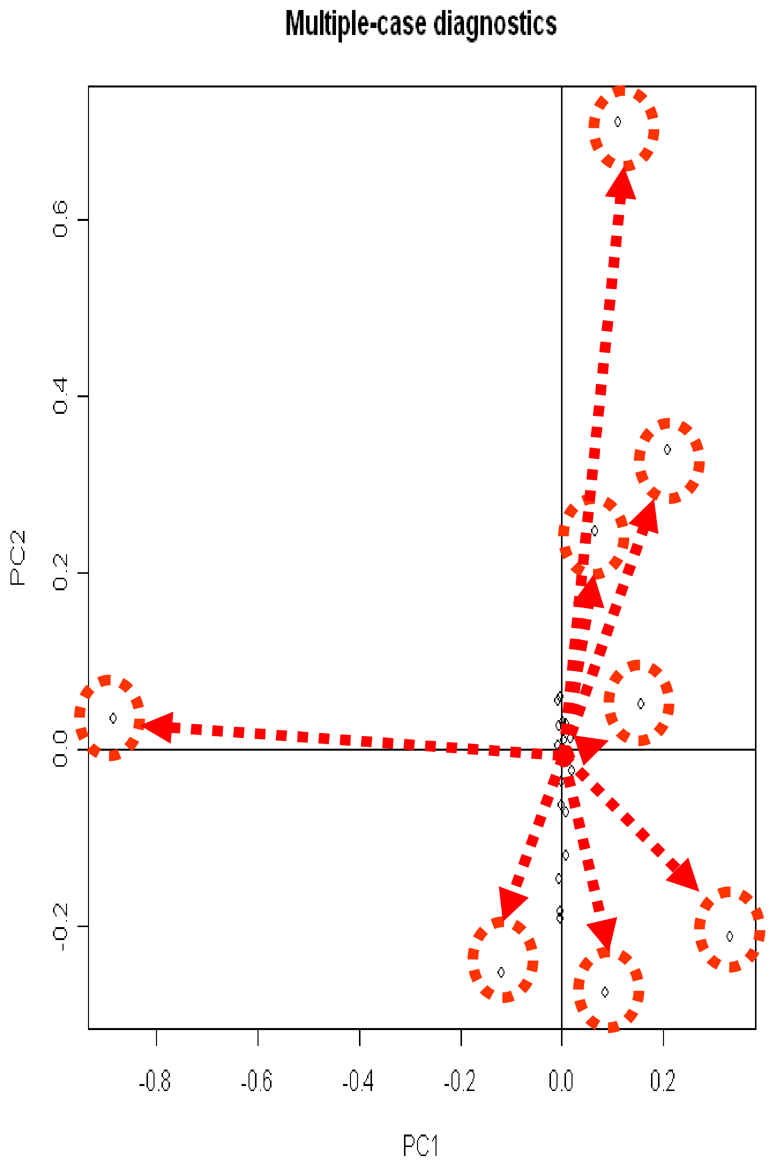
$$\left[[\mathbf{EIF}_{kg}] \left[\hat{\mathbf{acov}} \left(\mathbf{vech}\left(\hat{Q}_k \right) \right) \right]^{-1} [\mathbf{EIF}_{kg}]^{\mathrm{T}} - \lambda \mathbf{I} \right] \mathbf{h} = 0$$

2 Plot them in a lower dimensional space

3 Detect observations that have similar influence

4 Evaluate the values of

$$\hat{Z}_k^{g\,(J)} \left(J \in \{1, \ldots, n_g\} \right)$$



Outline

1 Introduction
2 Sensitivity analysis in linear subspace method
3 A multiple-case diagnostics with clustering
4 Numerical example
5 Concluding remarks

A multiple-case diagnostics with clustering

We can not always represent influence directions in low dimensions.

⇨ We solve the problem with clustering.

Eigenvalues of $\left[[\mathbf{EIF}_{kg}] \left[\hat{\mathbf{acov}} \left(\mathbf{vech} \left(\hat{Q}_k \right) \right) \right]^{-1} [\mathbf{EIF}_{kg}]^{\mathrm{T}} - \lambda \mathbf{I} \right]$:

$$\xi_1 \geq \xi_2 \geq \cdots \geq \xi_{n_g} \geq 0$$

A component of the eigenvectors : $\mathbf{a}_{ji} \left(j, i = 1, \ldots, n_g \right)$

Euclidean norm

$$d_{jj} = \sqrt{\sum_{i=1}^{n_g} \left(\mathbf{a}_{ji} - \mathbf{a}_{ji}\right)^{\mathrm{T}} \left(\mathbf{a}_{ji} - \mathbf{a}_{ji}\right)} \left(j, j = 1, \ldots, n_g\right).$$

Apply Ward's method to $D = \left(d_{jj}\right)$

Reduce the number of combinations of deleting observations based on the result

Outline

Numerical example

[Simulation]

Classes : Group1, Group2

Distribution : Multivariate normal distribution

Number of variables : 6

Number of observations : 10

Threshold of linear subspace method : 0.997

We compared the result of the proposed method with the result of deleting observations in all possible combinations.

Group1 :

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 105.857 & 127.241 & 0.037 & 1.030 & 18.244 & -12.352 \end{bmatrix}^{\mathbf{T}}$$

$$\Sigma_1 = \begin{pmatrix} 8676.883 & 789.092 & -0.384 & -21.393 & -525.731 & 245.875 \\ 789.092 & 1674.438 & -0.208 & -9.090 & 209.257 & -13.106 \\ -0.384 & -0.208 & 0.000 & 0.001 & 0.076 & -0.045 \\ -21.393 & -9.090 & 0.001 & 2.171 & 12.127 & -5.694 \\ -525.731 & 209.257 & 0.076 & 12.127 & 250.459 & -101.179 \\ 245.875 & -13.106 & -0.045 & -5.694 & -101.179 & 44.963 \end{pmatrix}$$

Data	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	3.251	74.731	0.04	2.868	19.71	−15.56
[2,]	−18.51	87.221	0.038	0.316	17.876	−13.239
[3,]	162.092	130.641	0.03	0.835	−2.196	−2.52
[4,]	120.625	175.7	0.033	3.239	39.097	−17.958
[5,]	70.893	104.692	0.064	−0.848	12.563	−10.372
[6,]	61.044	139.198	0.044	1.157	14.358	−10.993
[7,]	277.224	98.071	0.051	1.425	18.551	−15.139
[8,]	40.983	188.004	0.042	0.915	48.729	−23.67
[9,]	212.658	175.569	0.008	−1.482	−2.177	−1.233
[10,]	128.307	98.585	0.018	1.873	15.924	−12.834

Group2 :

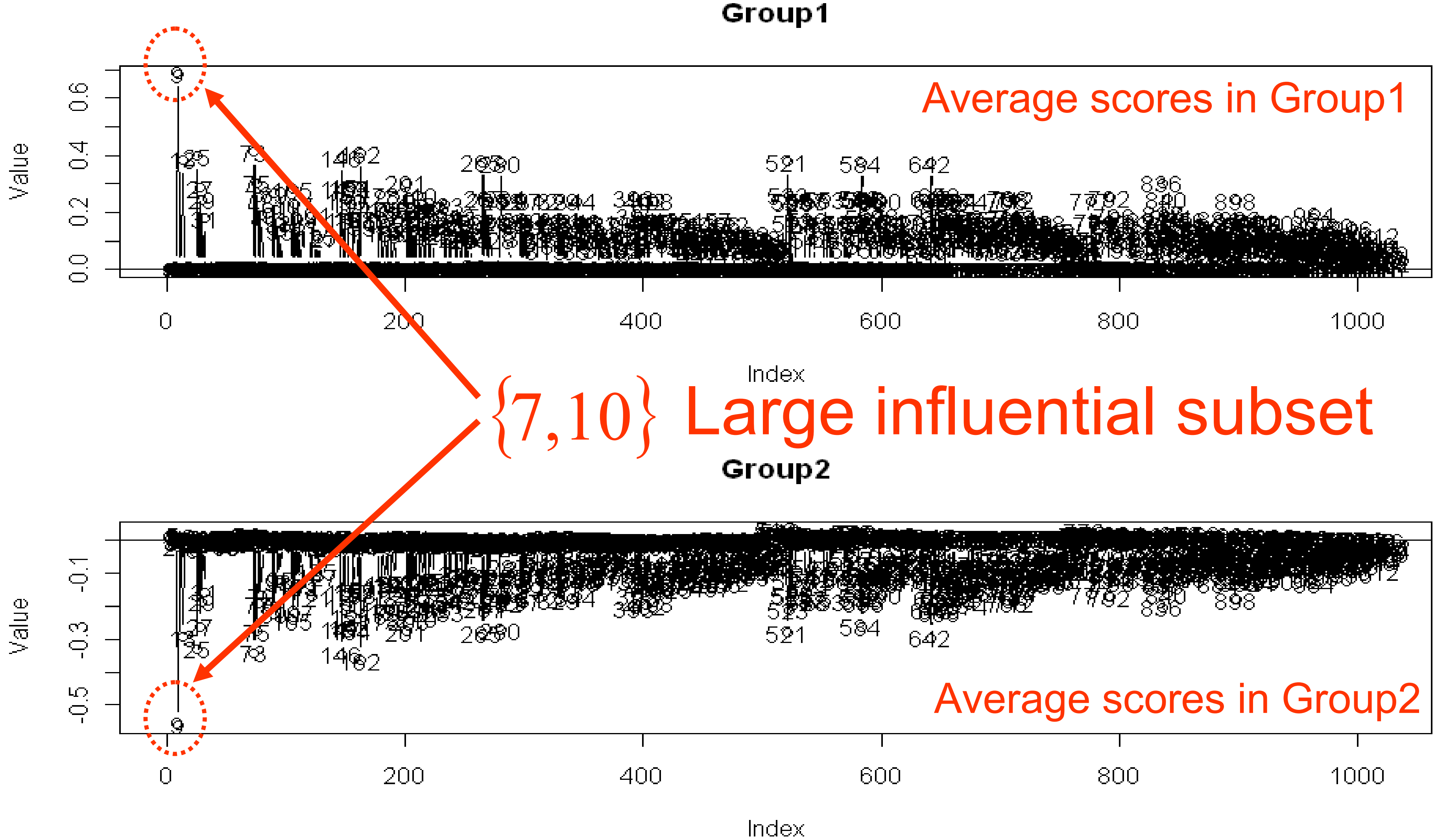
$$\boldsymbol{\mu}_2 = \begin{bmatrix} 97.492 & 131.578 & 0.000 & -0.089 & 1.325 & 1.314 \end{bmatrix}^{\mathrm{T}}$$

$$\Sigma_2 = \begin{pmatrix} 5736.272 & -84.440 & -1.018 & 17.550 & 145.029 & 77.281 \\ -84.440 & 1934.002 & -0.125 & 10.827 & 171.388 & 311.496 \\ -1.018 & -0.125 & 0.001 & 0.002 & -0.025 & 0.119 \\ 17.550 & 10.827 & 0.002 & 18.123 & 32.238 & 6.407 \\ 145.029 & 171.388 & -0.025 & 32.238 & 92.732 & 50.900 \\ 77.281 & 311.496 & 0.119 & 6.407 & 50.900 & 101.041 \end{pmatrix}$$

Data	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	90.704	96.139	−0.021	−0.622	−0.666	−3.766
[2,]	39.285	95.397	0.026	3.493	0.681	1.991
[3,]	31.19	243.481	−0.001	1.623	8.291	16.302
[4,]	−18.781	113.201	0.011	2.88	7.173	−4.303
[5,]	193.847	161.998	0.024	0.688	15.552	22.948
[6,]	122.154	110.92	0.024	4.865	7.591	−4.361
[7,]	174.834	111.309	0.007	−0.382	−7.5	−1.789
[8,]	48.081	122.082	0.024	−9.056	−15.391	−2.228
[9,]	87.225	122.684	−0.013	−5.637	−9.465	−4.696
[10,]	206.376	138.573	−0.082	1.261	6.988	−6.954

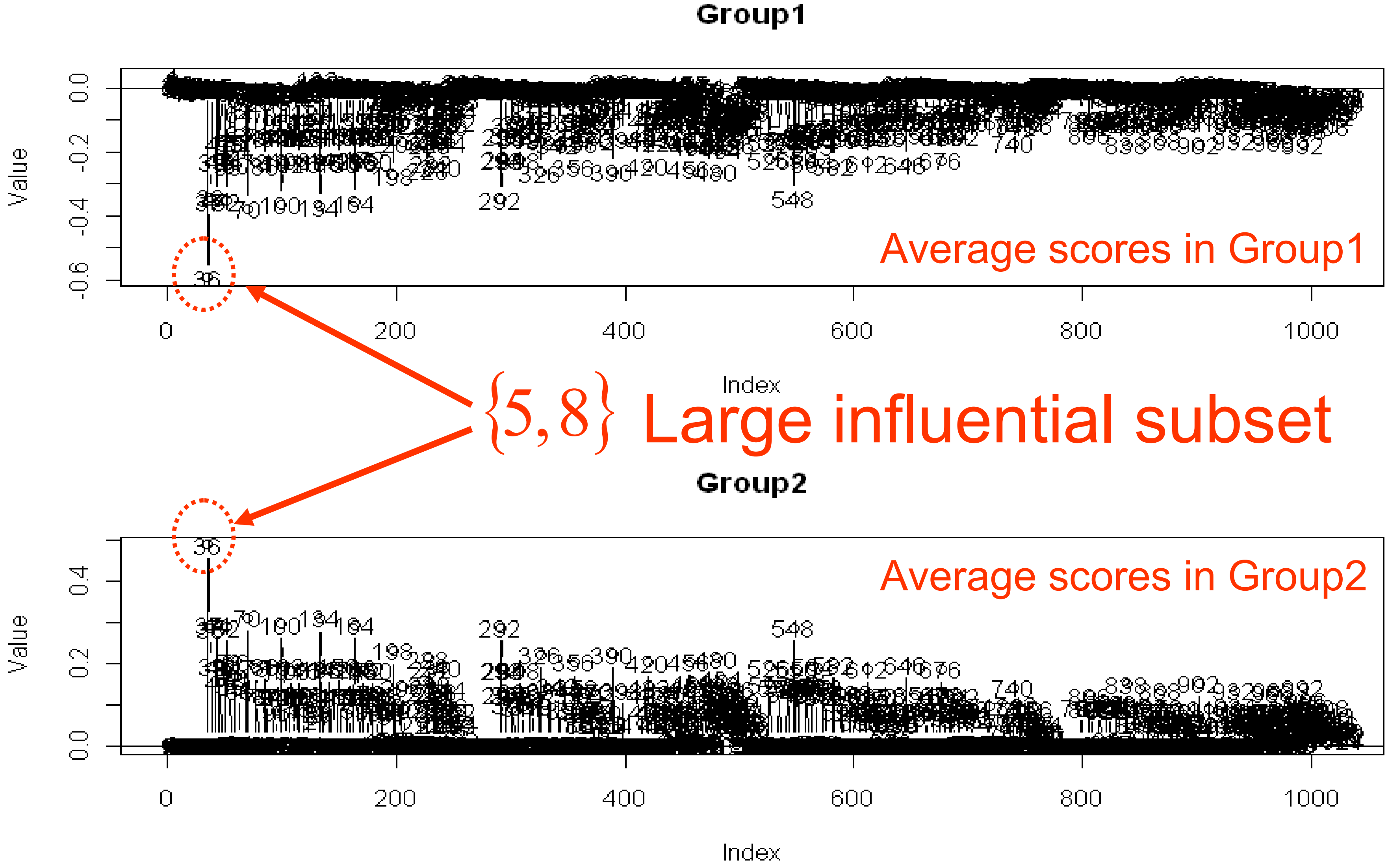
All possible combinations

$$\hat{Z}_k^{1(J)} \left(k = 1,2;\ J \in \{1,\ldots,n_1\}\right)$$



All possible combinations

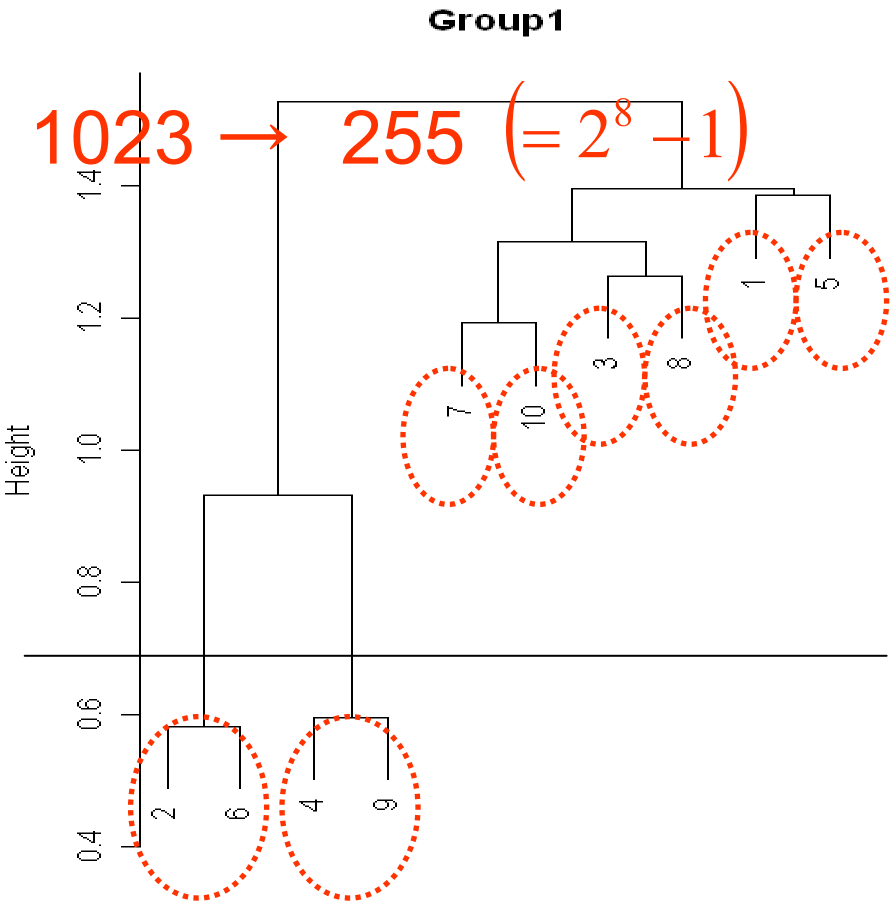
$$\hat{Z}_k^{2(J)} \ (k = 1,2; J \in \{1,\ldots,n_2\})$$



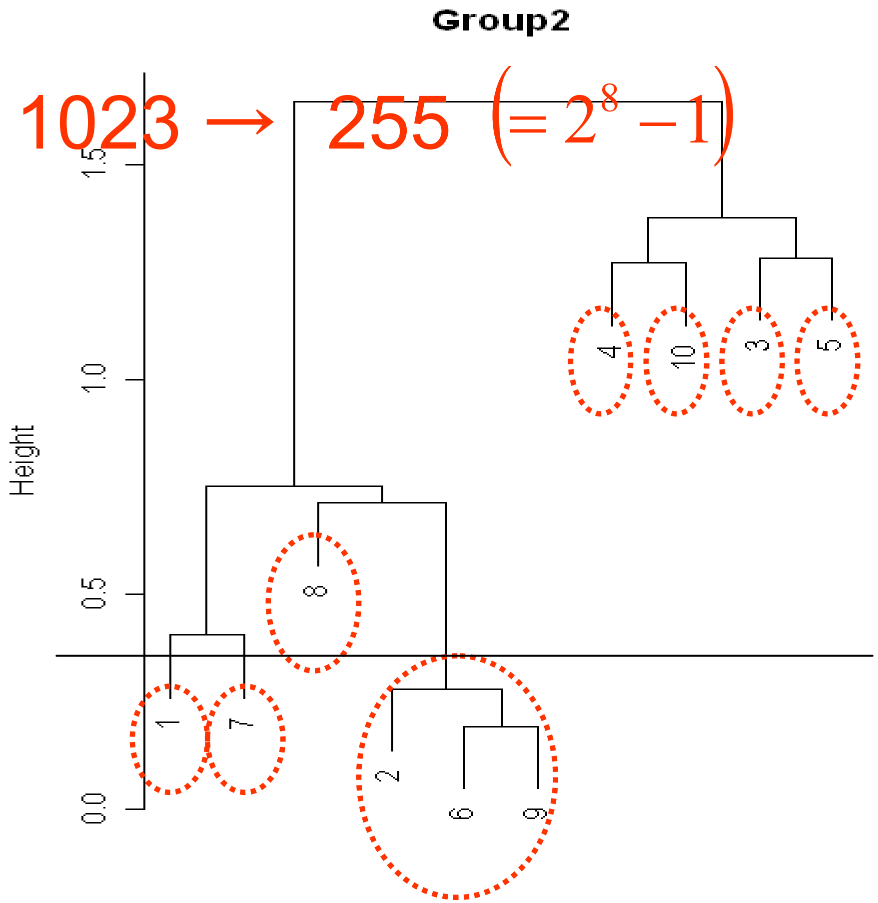
Proposed method

8 clusters　　　　8 clusters

$1023 \rightarrow 255 \ (= 2^8 - 1)$

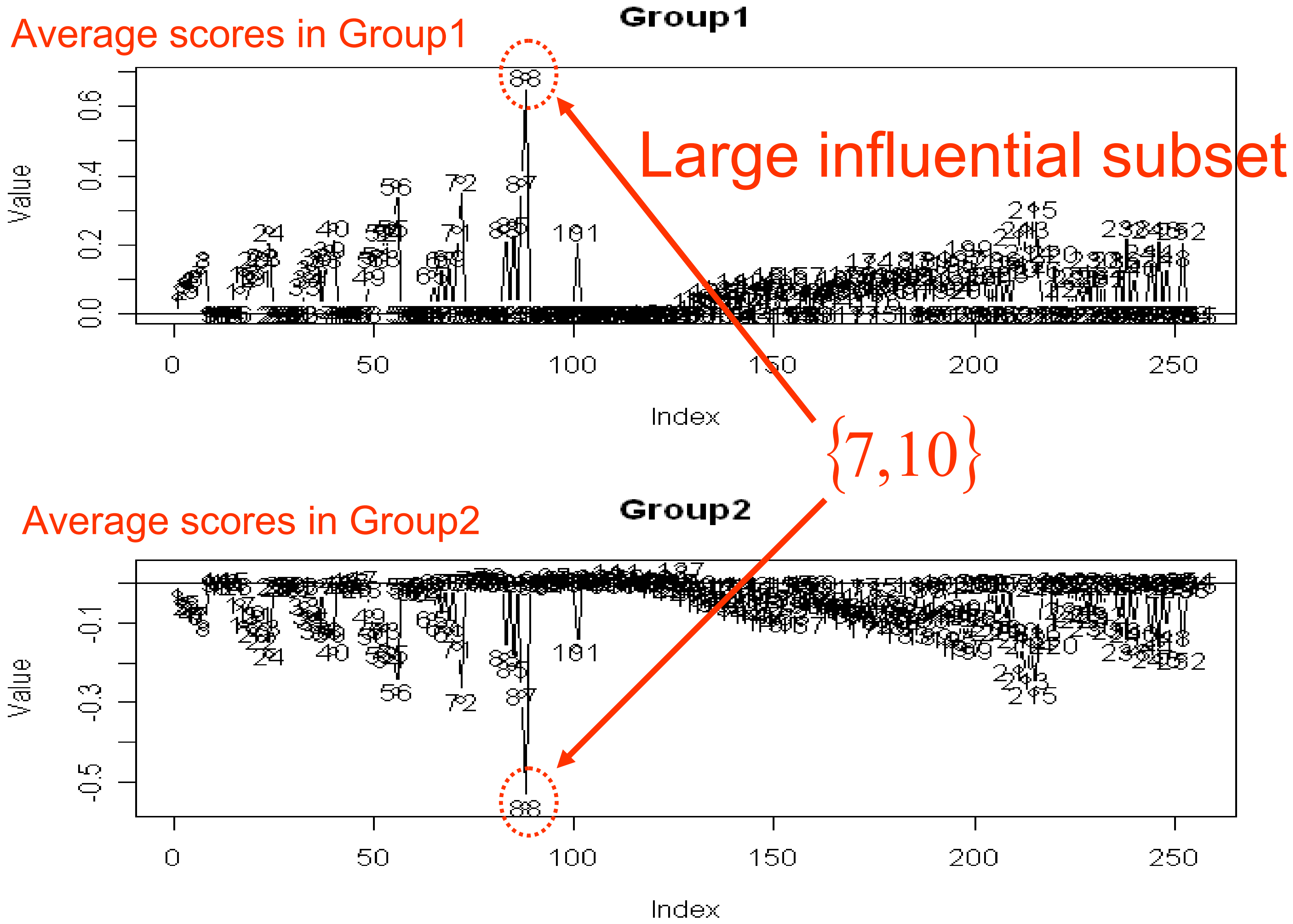


$1023 \rightarrow 255 \ (= 2^8 - 1)$

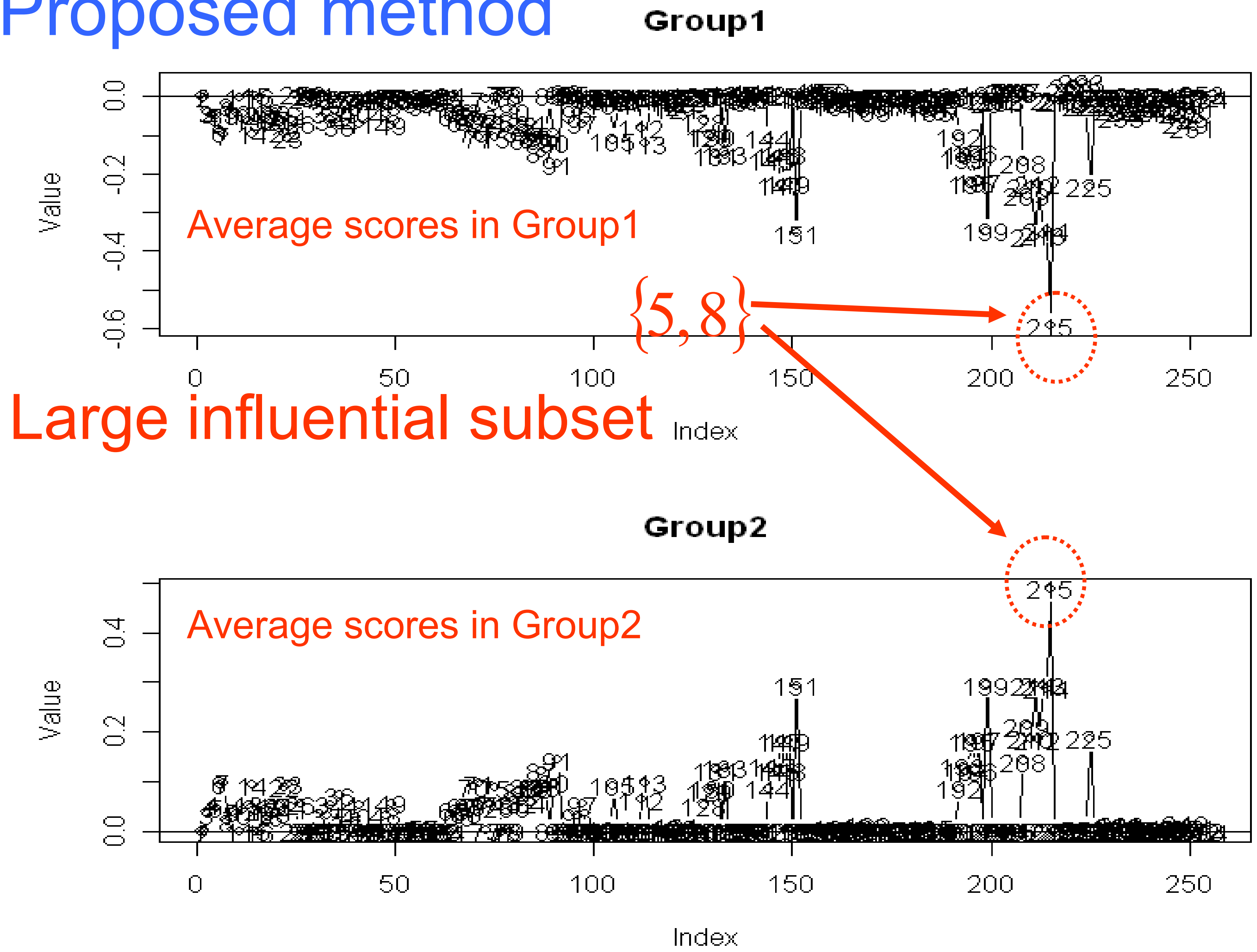


Clusters of influence directions

Proposed method



Proposed method



Case 2

[Simulation]

Classes : Group1, Group2

Distribution : Multivariate normal distribution

Number of variables : 6

Number of observations : 100

Threshold of linear subspace method : 0.997

Group1

$$\boldsymbol{\mu}_1 = [105.857 \quad 127.241 \quad 0.037 \quad 1.030 \quad 18.244 \quad -12.352]^T$$

$$\Sigma_1 = \begin{pmatrix} 8676.883 & 789.092 & -0.384 & -21.393 & -525.731 & 245.875 \\ 789.092 & 1674.438 & -0.208 & -9.090 & 209.257 & -13.106 \\ -0.384 & -0.208 & 0.000 & 0.001 & 0.076 & -0.045 \\ -21.393 & -9.090 & 0.001 & 2.171 & 12.127 & -5.694 \\ -525.731 & 209.257 & 0.076 & 12.127 & 250.459 & -101.179 \\ 245.875 & -13.106 & -0.045 & -5.694 & -101.179 & 44.963 \end{pmatrix}$$

Setting up 3 outliers that have similar influence direction

$$\{42, 93, 94\}$$

Group2

$$\boldsymbol{\mu}_2 = [97.492 \quad 131.578 \quad 0.000 \quad -0.089 \quad 1.325 \quad 1.314]^T$$

$$\Sigma_2 = \begin{pmatrix} 5736.272 & -84.440 & -1.018 & 17.550 & 145.029 & 77.281 \\ -84.440 & 1934.002 & -0.125 & 10.827 & 171.388 & 311.496 \\ -1.018 & -0.125 & 0.001 & 0.002 & -0.025 & 0.119 \\ 17.550 & 10.827 & 0.002 & 18.123 & 32.238 & 6.407 \\ 145.029 & 171.388 & -0.025 & 32.238 & 92.732 & 50.900 \\ 77.281 & 311.496 & 0.119 & 6.407 & 50.900 & 101.041 \end{pmatrix}$$

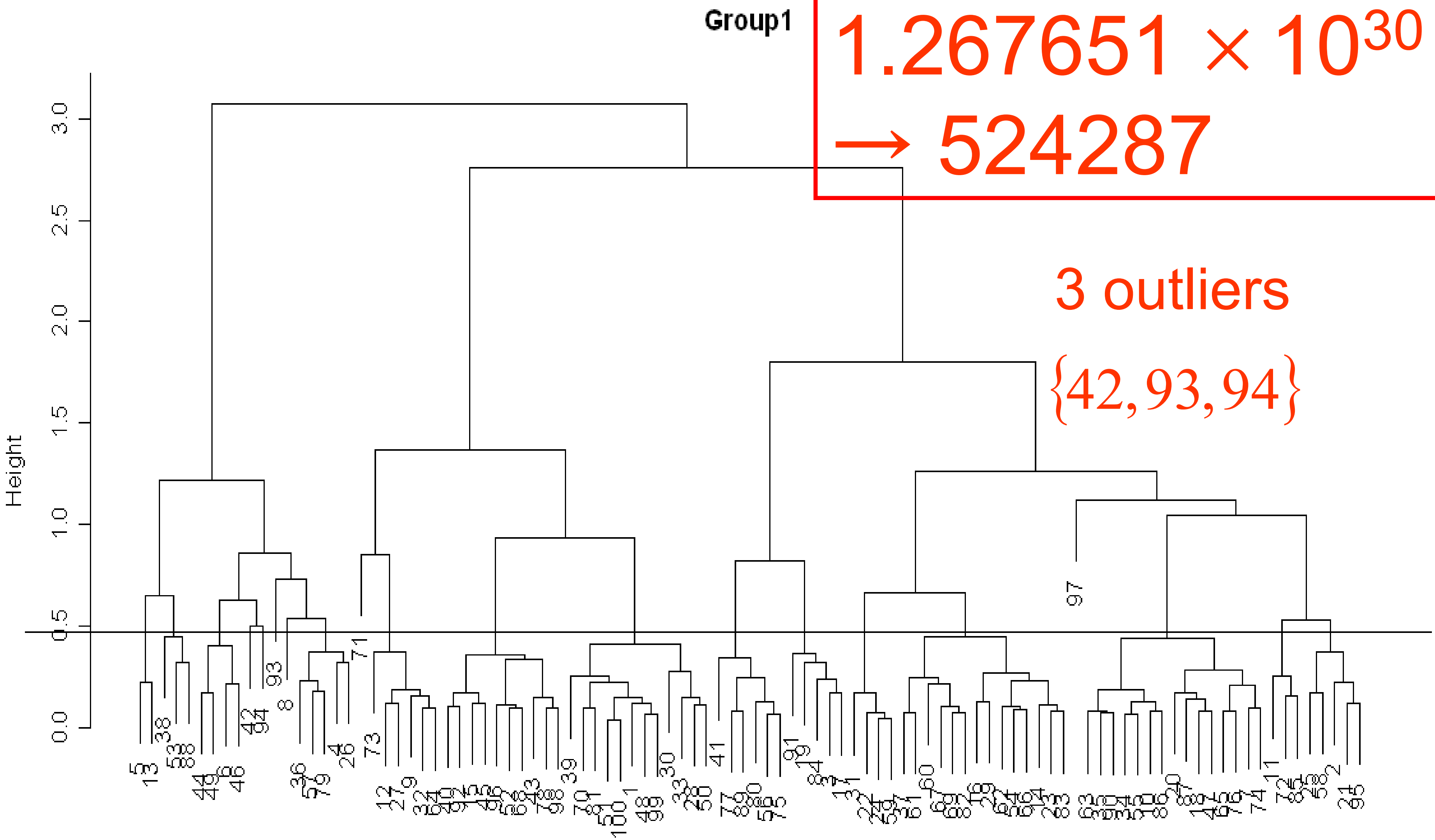
Our Multiple-case diagnostics in Group1

19 clusters

1.267651×10^{30}
$\rightarrow 524287$

3 outliers

$\{42, 93, 94\}$



Outline

1 Introduction
2 Sensitivity analysis in linear subspace method
3 A multiple-case diagnostics with clustering
4 Numerical example
5 Concluding remarks

Concluding remarks

We proposed a multiple-case diagnostics with clustering in linear subspace method.

In simulation study, we could confirm the availability of our method.

In the future, we would like to show the results of some simulation studies and their Monte Carlo simulations.

Thank you for your attention!