

## Sensitivity Analysis for Multiple Similarity Method and its Application

Kuniyoshi, Hayashi

*Graduate School of Information Sciences and Technology, Hokkaido University*

*N14W9, Kita-ku*

*Sapporo, Hokkaido, 0600814, JAPAN*

*E-mail: k-hayashi@iic.hokudai.ac.jp*

Yutaka, Tanaka

*Institute of Statistical Mathematics*

*10-3, Midori-cho*

*Tachikawa, Tokyo, 1908562, JAPAN*

*E-mail: ytanaka@ems.okayama-u.ac.jp*

### Introduction

Sensitivity analysis based on influence functions has been developed for many multivariate statistical methods. To evaluate the influence of observations and detect outliers in linear discriminant analysis, the influence functions have been derived by Campbell (1978), Huang *et al.* (2007) and others and these influence functions were used for detecting influential observations to the result of analysis.

In the field of pattern recognition, there have been developed several kinds of discriminant analysis. They include a group of methods called by linear subspace methods (Oja, 1983) such as CLAss Featuring Information Compression or CLAFIC method (Watanabe, 1967) and Multiple Similarity Method or MSM (Iijima, 1973; Omachi and Aso, 2000). The basic idea of the subspace methods is to find a low dimensional subspace which represents the distribution of each class in the feature space and evaluate in which subspace an observation or unknown feature vector can be approximated well.

For these methods of discriminant analysis, it would be important to develop methods of sensitivity analysis or statistical diagnostics as it is for classical discriminant analysis. From this perspective, Hayashi *et al.* (2010) studied sensitivity analysis for CLAFIC method and proposed a method of multiple-case as well as single-case diagnostics.

In this paper, we study sensitivity analysis for MSM. At first, we introduce a discriminant score to measure the goodness of classification, then derive sample and empirical influence functions for a statistics which plays an important role in the discriminant score and finally propose a procedure of sensitivity analysis for MSM using these influence functions with a numerical example to illustrate the effectiveness of the proposed method.

### Multiple Similarity Method (MSM)

Let  $\mathbf{x}_i^k$  ( $i = 1, \dots, n_k$ ;  $k = 1, \dots, K$ ) be the  $i$ -th  $p$  dimensional observations in the  $k$ -th class of the training sample. Then, the autocorrelation matrix in the  $k$ -th class is  $1/n_k \sum_{i=1}^{n_k} \mathbf{x}_i^k \mathbf{x}_i^{kT}$  ( $= \hat{G}_k$ ). Using the eigenvalues  $\hat{\lambda}_i^k$  and the eigenvectors  $\hat{\mathbf{u}}_i^k$ , multiple similarity or transformation matrix from original feature space to the subspace is define as

$$(1) \quad \hat{M}_k = \sum_{s=1}^{p_k} \frac{\sqrt{\hat{\lambda}_s^k}}{\sqrt{\hat{\lambda}_1^k}} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{kT}, \quad (1 \leq p_k \leq p),$$

where  $p_k$  is equal to the minimum value  $m$  satisfying  $\tau \leq \frac{\sum_{s=1}^m \hat{\lambda}_s^k}{\sum_{s=1}^p \hat{\lambda}_s^k}$  ( $1 \leq m \leq p$ ). When an input

datum comes as a test observation  $\mathbf{x}$ , we project  $\mathbf{x}$  to  $\hat{M}_k \mathbf{x}$  and calculate the projected norm onto each class,  $\|\hat{M}_k \mathbf{x}\|$ . Next, we calculate the squared value each class with the projection norm and classify  $\mathbf{x}$  into the class that gives the maximum squared projection norm.

**Discriminant score and its average**

We define a discriminant score of  $\mathbf{x}_i^k$  as

$$(2) \quad \hat{z}_i^k = \mathbf{x}_i^{kT} \hat{Q}_k \mathbf{x}_i^k, \quad (1 \leq i \leq n_k),$$

where

$$(3) \quad \hat{Q}_k = \frac{K}{K-1} \left( \hat{M}_k - \frac{1}{K} \sum_{\ell=1}^K \hat{M}_\ell \right).$$

In addition, we calculate an average of the discriminant scores as follows.

$$(4) \quad \hat{Z}^k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{z}_i^k.$$

If the value of  $\hat{Z}^k$  is large, we can understand that the  $k$ -th class is separated well from other classes.

**Influence functions for  $\hat{Q}_k$**

In MSM,  $\hat{Q}_k$  in equation (3) is an important statistics. This statistics appears in not only equation (2) but also equation (4). Then, we particularly focus on  $\hat{Q}_k$ . If the sample influence function has a good approximation for the empirical influence function, to detect the relationship of influence patterns of observations in multiple-case diagnostics, we can use the additive property of empirical influence function (Tanaka, 1994). In this section, we derive the sample and empirical influence functions.

The sample influence function for  $\hat{Q}_k$  at a point of  $\mathbf{x}_j^g$  is

$$(5) \quad \text{SIF}(\mathbf{x}_j^g; \hat{Q}_k) \equiv \hat{Q}_k^{(1)g(-j)} = -(n_g - 1) \cdot (\hat{Q}_k^{g(-j)} - \hat{Q}_k),$$

where  $g = 1, \dots, K$  and  $j = 1, \dots, n_g$ . The superscript notation  $g(-j)$  represents deleting  $j$ -th observation in  $g$ -th class.

The empirical influence function for  $\hat{Q}_k$  at a point of  $\mathbf{x}_j^g$  is

$$(6) \quad \text{EIF}(\mathbf{x}_j^g; \hat{Q}_k) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{Q}_k^{gj} - \hat{Q}_k}{\varepsilon} \equiv \hat{Q}_k^{(1)gj} = \gamma \cdot \left\{ \sum_{s=1}^{p_g} \sum_{t=p_g+1}^p \frac{\hat{\lambda}_s^g}{\hat{\lambda}_1^g} (\hat{\lambda}_s^g - \hat{\lambda}_t^g)^{-1} \hat{a}_{st}^{g(1)} (\hat{\mathbf{u}}_s^g \hat{\mathbf{u}}_t^{gT} + \hat{\mathbf{u}}_t^g \hat{\mathbf{u}}_s^{gT}) \right. \\ \left. + \sum_{s=1}^{p_g} \sum_{t=1}^{p_g} \frac{\hat{a}_{st}^{g(1)}}{\hat{\lambda}_1^g} \hat{\mathbf{u}}_s^g \hat{\mathbf{u}}_t^{gT} - \sum_{s=1}^{p_g} \frac{\hat{\lambda}_s^g \hat{\lambda}_1^{g(1)}}{\hat{\lambda}_1^{g2}} \hat{\mathbf{u}}_s^g \hat{\mathbf{u}}_s^{gT} \right\},$$

where  $\hat{a}_{st}^{g(1)}$  is equal to  $\hat{\mathbf{u}}_s^{gT} (\hat{\mathbf{x}}_j^g \hat{\mathbf{x}}_j^{gT} - \hat{G}_g) \hat{\mathbf{u}}_t^g$  ( $j = 1, \dots, n_g$ ) and  $\hat{\lambda}_1^{g(1)}$  is equivalent to  $\hat{a}_{11}^{g(1)}$ . The superscript notation  $gj$  means the derivation of  $\hat{Q}_k$  at a point of  $\mathbf{x}_j^g$ . In the case of perturbing  $\mathbf{x}_j^g$  ( $g = k$ ),  $\gamma$  is equal to 1. In other cases ( $g \neq k$ ),  $\gamma$  is equivalent to  $-\frac{1}{K-1}$ .

**Sensitivity analysis for MSM**

In this section, we develop sensitivity analysis for MSM including a single-case and multiple-case diagnostics based on Tanaka (1994).

**Single-case diagnostics**

Since  $SIF(\mathbf{x}_j^g; \hat{Q}_k)$  and  $EIF(\mathbf{x}_j^g; \hat{Q}_k)$  are matrices, to evaluate the influence of a sample observation for  $\hat{Q}_k$ , we summarize them into scalar measures  $\hat{Z}_k^{g(-j)}$  and  $\hat{Z}_k^{gj}$ ,

$$(7) \quad \hat{Z}_k^{g(-j)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^k \text{T} \hat{Q}_k^{(1)g(-j)} \mathbf{x}_i^k$$

and

$$(8) \quad \hat{Z}_k^{gj} = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_i^k \text{T} \hat{Q}_k^{(1)gj} \mathbf{x}_i^k.$$

If the value of  $\hat{Z}_k^{g(-j)}$  or  $\hat{Z}_k^{gj}$  is large, we can confirm that a point of  $\mathbf{x}_j^g$  has large influence.

**Multiple-case diagnostics**

In this study, we evaluate the influence of multiple observations for  $\hat{Q}_k$  based on the discussion about Cook (1986), Tanaka (1994), Tanaka *et al.* (1999) and Tanaka *et al.* (2003). Here, we determine a target statistics as  $\text{vech}(\hat{Q}_k)$  by considering the symmetric property of  $\hat{Q}_k$ . We denote the unperturbed weights for  $n_g$  observations as  $\mathbf{w}_0^g = (1, 1, \dots, 1)^T$  and also represent the weights for the perturbed observations as  $\mathbf{w}^g$ . In the framework of Cook’s diagnostics (1986), to evaluate the influence of observations, the change from  $\text{vech}(\hat{Q}_k)$  that is estimated by the log likelihood function for the unperturbed case,  $L(\text{vech}(Q_k)|\mathbf{w}_0^g)$  to  $\text{vech}(\hat{Q}_{k\mathbf{w}^g})$  that is estimated by the log likelihood function for the perturbed case,  $L(\text{vech}(Q_k)|\mathbf{w}^g)$  is measured as

$$(9) \quad D(\mathbf{w}^g) = 2 \left[ L(\text{vech}(\hat{Q}_k)|\mathbf{w}_0^g) - L(\text{vech}(\hat{Q}_{k\mathbf{w}^g})|\mathbf{w}_0^g) \right].$$

Then, the change along a line  $\mathbf{w}^g = \mathbf{w}_0^g + t\mathbf{h}$  ( $\|\mathbf{h}\| = 1$ ) plays an important role. Cook (1986) investigated the direction that had the largest curvature at  $\mathbf{w}_0^g$ .  $D(\mathbf{w}^g)$  can be rewritten as

$$(10) \quad C_{\mathbf{h}} = \mathbf{h}^T \left[ \frac{\partial \text{vech}(\hat{Q}_{k\mathbf{w}^g})^T}{\partial \mathbf{w}^g} \right] \left[ -\frac{\partial^2 L}{\partial \text{vech}(Q_k) \partial \text{vech}(Q_k)^T} \right] \left[ \frac{\partial \text{vech}(\hat{Q}_{k\mathbf{w}^g})}{\partial \mathbf{w}^{gT}} \right] \mathbf{h},$$

where  $[\partial \text{vech}(\hat{Q}_{k\mathbf{w}^g}) / \partial \mathbf{w}^g]$  and  $[-\partial^2 L / \partial \text{vech}(Q_k) \partial \text{vech}(Q_k)^T]$  are evaluated at  $\mathbf{w}^g = \mathbf{w}_0^g$  and  $\text{vech}(Q_k) = \text{vech}(\hat{Q}_k)$ , respectively (Tanaka, 1994). To search the most influential direction, we have to maximize  $C_{\mathbf{h}}$  under  $\|\mathbf{h}\| = 1$ . In this regard,  $[\partial \text{vech}(\hat{Q}_{k\mathbf{w}^g}) / \partial \mathbf{w}^g]$  is  $(1/n_g) \cdot \left\{ EIF(\mathbf{x}_j^g; \text{vech}(\hat{Q}_k)) \right\}$ . We denote  $\left\{ EIF(\mathbf{x}_j^g; \text{vech}(\hat{Q}_k)) \right\}$  as  $[EIF_g]$ . In addition, by an asymptotically normally and efficient properties in an ML estimator, equation (10) can be written as  $C_{\mathbf{h}} = \mathbf{h}^T [EIF_g] \left[ \widehat{\text{acov}}(\text{vech}(\hat{Q}_k)) \right]^{-1} [EIF_g]^T \mathbf{h}$ . Then, the most influential direction is calculated as the eigenvector associated with the largest eigenvalue in  $[EIF_g] [\widehat{\text{acov}}(\text{vech}(\hat{Q}_k))]^{-1} [EIF_g]^T$ . In the estimation of  $\widehat{\text{acov}}(\text{vech}(\hat{Q}_k))$ , we use jackknife method as follows.

$$\widehat{\text{acov}}(\text{vech}(\hat{Q}_k)) \cong V_{\text{JACK} / k} =$$

$$(11) \quad \frac{R-1}{R} \sum_{i=1}^R \left[ \text{vech}(\hat{Q}_k)_{R-1,i} - \frac{1}{R} \sum_{j=1}^R \text{vech}(\hat{Q}_k)_{R-1,j} \right] \left[ \text{vech}(\hat{Q}_k)_{R-1,i} - \frac{1}{R} \sum_{j=1}^R \text{vech}(\hat{Q}_k)_{R-1,j} \right]^T,$$

where  $R$  is the number of total sample size.

From equation (6), we can understand that the influence of  $\mathbf{x}_j^k$  for  $\hat{Q}_k$  is  $-(K - 1)$  times that of  $\mathbf{x}_j^g$  ( $g \neq k$ ) for  $\hat{Q}_k$ . Then, we only focus on the influence of  $\mathbf{x}_j^k$  for  $\hat{Q}_k$  to search the similar influence patterns of observations in each class. To get the influence directions of observations for  $\hat{Q}_k$ , we calculate the following eigenvalue problem.

$$(12) \quad \left[ [\text{EIF}_k] \left[ \widehat{\text{acov}}(\text{vech}(\hat{Q}_k)) \right]^{-1} [\text{EIF}_k]^T - \lambda \mathbf{I} \right] \mathbf{h} = 0.$$

With our proposed method in multiple-case diagnostics, we can visually find the influence patters of multiple observations with a few dimensions.

### Application

We show the effectiveness of our sensitivity analysis through a numerical example. Here, we deal with the identification problem of 5 classes. The number of dimensions and that of observations in each class are 100 and 30, respectively. Datasets of Classes 1, 4 and 5 are generated assuming 3-contaminated multivariate normal distributions with different parameters, and those of Classes 2 and 3 are generated assuming 2-contaminated multivariate normal distribution and a multivariate normal distribution, respectively. To contaminate outlying observations, we generated the 23rd and 24th observations of Class 5 using the parameter of Class 4 instead of Class 5.

Under the above situation, we firstly set that the norms of all the training and test observations were normalized to one. We determined the value of  $\tau$  as 0.9973 based on leave-one out cross validation and developed a classifier with MSM. The results of the training and test datasets in each class are shown in Table 1. From these results, we could understand that the observations in Class 4 and Class 5 were not discriminated well. In this situation, we firstly applied our single-case diagnostics.

When we perturbed the 23rd and 24th observations in Class 5, the scalar influence measures by the sample influence function showed the change that the observations in Class 4 were discriminated well than Class 5 (Figure 1). To confirm the relationship of the sample influence function and empirical influence function, we plotted  $\sum_{k=1}^5 \hat{Z}_k^{5(-j)}$  and  $\sum_{k=1}^5 \hat{Z}_k^{5j}$  ( $j = 1, \dots, 30$ ) in a horizontal  $x$ -axis and vertical  $y$ -axis, respectively (Figure 2). From Figure 2, we could see that the sample influence function was nearly approximated with the empirical influence function. Figure 3 shows the scatterplots of PC scores of influence functions of observations in Class 5 for PC1 and PC2. From Figure 3, we could see that they had the same influence. We finally deleted them in training data at a time and redeveloped the classifier so that we got Table 2.

### Concluding remarks

In this paper, we defined a discriminant score for MSM and we derived the influence functions. With these influence functions, we proposed sensitivity analysis for MSM and illustrated the effectiveness of our method through a numerical example.

Table 1: Results of MSM applied to training and test datasets

	Class1	Class2	Class3	Class4	Class5		Class1	Class2	Class3	Class4	Class5
Class1	30	0	0	0	0	Class1	300	0	0	0	0
Class2	0	30	0	0	0	Class2	0	300	0	0	0
Class3	0	0	30	0	0	Class3	0	0	300	0	0
Class4	0	0	0	24	6	Class4	0	0	0	230	70
Class5	0	0	0	0	30	Class5	0	0	0	0	300

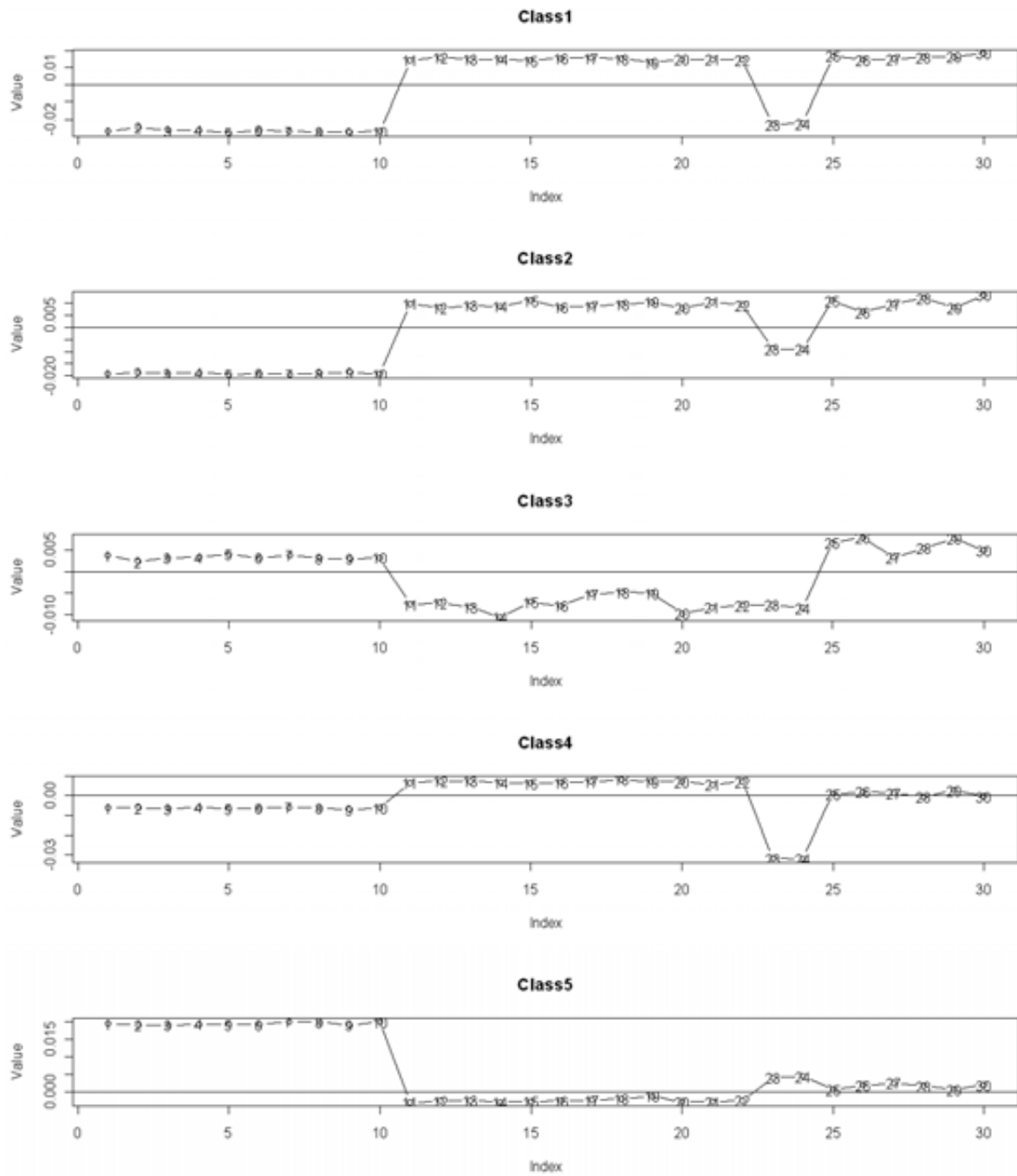


Figure 1: Indexplot of the average of discriminant scores for the observations in each class

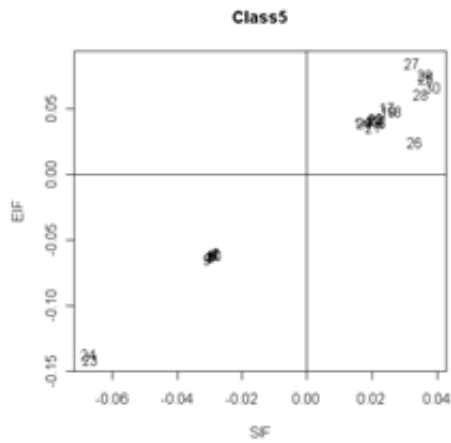


Figure 2: SIF vs EIF

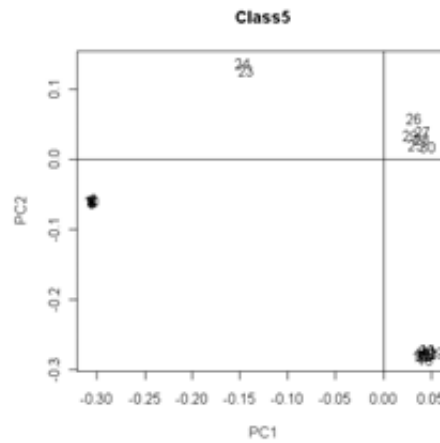


Figure 3: Influence patterns

Table 2: Results of training and test data after diagnostics

	Class1	Class2	Class3	Class4	Class5
Class1	30	0	0	0	0
Class2	0	30	0	0	0
Class3	0	0	30	0	0
Class4	0	0	0	30	0
Class5	0	0	0	2	28

	Class1	Class2	Class3	Class4	Class5
Class1	300	0	0	0	0
Class2	0	300	0	0	0
Class3	0	0	300	0	0
Class4	0	0	0	300	0
Class5	0	0	0	0	300

## REFERENCES

- Campbell, N. A. (1978). The influence function as an aid in outlier detection in discriminant analysis, *Applied Statistics*, **27**, 3, 251-258.
- Cook, R. D. (1986). Assessment of local influence, *Journal of the Royal Statistical Society, B*, **48**, 133-169.
- Hayashi, K., Minami, H. and Mizuta, M. (2010). On Multiple-Case Diagnostics in Linear Subspace Method, *Proceedings in Computational Statistics 2010*, Springer-Verlag Berlin Heidelberg 2010, 493-500.
- Huang, Y., Kao, T. L. and Wang, T. H. (2007). Influence functions and local influence in linear discriminant analysis. *Computational Statistics and Data Analysis*, **51**, 3844-3861.
- Iijima, T., Genchi, H. and Mori, K. (1973). A theory of character recognition by pattern matching method, *Proceedings of the 1st International Joint Conference on Pattern Recognition*, 50-56.
- Oja, E. (1983). *Subspace methods of pattern recognition*, Wiley and Sons, New York.
- Omachi, S. and Aso, H. (2000). A qualitative adaption of subspace method for character recognition, *Systems and Computers in Japan*, **31**, 12.
- Tanaka, Y. (1994). Recent advance in sensitivity analysis in multivariate statistical analysis, *Journal of the Japanese Society of Computational Statistics*, **7**, 1-25.
- Tanaka, Y. and Zhang, F. (1999). R-mode and Q-mode influence analyses in statistical modelling: relationship between influence function approach and local influence approach, *Computational Statistics and Data Analysis*, **32**, 197-218.
- Tanaka, Y., Zhang, F. and Mori, Y. (2003). Local influence in principal component analysis: relationship between the local influence and influence function approaches revisited, *Computational Statistics and Data Analysis*, **44**, 143-160.
- Watanabe, S., Lambert, P. F., Kulikowski, C. A., Buxton, J. L. and Walker, R. (1967). Evaluation and selection of variables in pattern recognition, *Computer and Information Sciences II*, Academic Press.