Perturbation Analysis for Similarity Based on Entropy in a Linear Subspace Method

Kuniyoshi Hayashi[1, 5], Fumio Ishioka[2, 5], Hiroshi Suito[3, 5], Koji Kurihara[4, 5]
[1] Graduate School of Environmental and Life Science, Okayama University, Okayama, JAPAN
[2] School of Law, Okayama University, Okayama, JAPAN
[3] Graduate School of Environmental and Life Science, Okayama University, Okayama, JAPAN
[4] Graduate School of Environmental and Life Science, Okayama University, Okayama, JAPAN
[5] CREST, the Japan Science and Technology Agency
[6] Corresponding author: Kuniyoshi Hayashi, e-mail: k-hayashi@ems.okayama-u.ac.jp

Abstracts

In the field of pattern recognition, there are many discriminant methods similar to statistical discriminant methods. However, statistical quantitative assessment studies for these classifiers have not been as advanced as one might expect. In statistics, an assessment method based on the influence function was proposed, and its approach was applied to various discriminant methods such as linear and quadratic ones. If we use the influence function in the assessment for statistical discriminant methods, we can detect large influential training samples for prediction accuracy and can also enhance the performance of classifiers on the basis of its finding. Till now, we have mainly focused on the class-featuring information compression (CLAFIC) method, which is a discriminant method in pattern recognition, and we have developed a diagnostic for this method by using the influence function. In this article, we particularly focus on the CLAFIC method for a similarity based on the entropy function. For this method, we perturb the average of the discriminant scores in each class at a training sample and calculate the influence functions for the statistics to evaluate the influence of the training sample for prediction accuracy. In addition, we suggest the possibility of multiple-case diagnostics for the CLAFIC method by using additivity of the empirical influence function.

Keywords: CLAFIC method, diagnostics, influence function, pattern recognition

1. Introduction

Recently, in many fields, we have been able to easily obtain a large-sized dataset from high-performance measurement hardware. When we analyze such large datasets to acquire findings for a decision process, we must reduce the size of the dataset, preserving its essential and informative structure with respect to memory size and calculation time. In face and speech recognition, we encounter a large-sized dataset, and therefore, we must reduce it to a manageable size to conduct preprocessing procedures. In the field of pattern recognition, a compression and classification method similar to principal component analysis and Karhunen-Loeve expansion was proposed by Watanabe et al. (1967), and it has been applied to certain discriminant problems. This method is called the class-featuring information compression (CLAFIC) method. By using the CLAFIC method, we can simultaneously and quickly perform data compression and discrimination. This method has been successfully used for a limited pattern recognition problem. However, since the kernel method was applied to the CLAFIC method (Maeda and Murase, 1999), the status of the discriminant method has dramatically risen, equaling that of support vector machine (SVM). Therefore, extensions of the CLAFIC method have been used in pattern recognition problems (e.g., Gunal and Edizkan, 2008). The CLAFIC method possesses two excellent characteristics relative to SVM. One is fast processing in huge multi-class discriminant problems and the other is a possibility of a statistical quantitative assessment for the prediction accuracy of a target classifier. In particularly, we have focused on the second characteristic and have proposed sensitivity analysis for the CLAFIC method on the basis of the influence function of Hampel (1974) and Hampel et al. (1986) with reference to Tanaka (1994). Moreover, we have considered sensitivity analysis for an extension of the CLAFIC method (Hayashi and Tanaka, 2011). According to Watanabe et al. (1967, pp. 91-122), Watanabe (1969), Watanabe (1970, pp. 63-111), and Oja (1983), there are two representative similarities in the CLAFIC method. One is a similarity composed of the projection norm calculated by projecting an input pattern vector into the subspace in each class. The other is a similarity by the entropy function based on the projection norm of an input pattern vector into the subspace in each class. In this paper, we focus on the CLAFIC method for a similarity based on the entropy function of Shannon (1948). Here we calculate a discriminant score for the method in each class and its average. In addition, we perturb these important statistics at a training sample in each class and calculate the influence functions for them. Finally, we describe a diagnostics of the CLAFIC method on the basis of the entropy function.

2. Discriminant rule by similarity based on entropy function

We denote the number of the dimensions of the pattern vector by p and assume the number of classes to be K. In addition, we denote the i-th training sample in the k-th class as $\mathbf{x}_i^k \in \mathbf{R}^p \ (i = 1,\ldots,n_k; k = 1,\ldots,K)$, where n_k is the number of the training samples in the k-th class. We normalize the norm of $\mathbf{x}_i^k$ into one and calculate the autocorrelation matrix in each class as $\hat{G}_k = (1/n_k)\sum_{i=1}^{n_k} \mathbf{x}_i^k \mathbf{x}_i^{k\mathrm{T}} \ (k = 1,\ldots,K)$. Next we solve the eigenvalue problem for $\hat{G}_k$. We denote the p eigenvalues as $\hat{\lambda}_1^k > \cdots > \hat{\lambda}_p^k \geq 0$. Moreover, we represent the normalized eigenvector corresponding to $\hat{\lambda}_s^k$ as $\hat{\mathbf{u}}_s^k \ (s = 1,\ldots,p)$. Here we assume that $\mathbf{x}$ is an input vector and calculate the square of the value of projection norm $\mathbf{x}^{\mathrm{T}}\sum_{s=1}^{p} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}} \mathbf{x}$. Now $\mathbf{x}^{\mathrm{T}} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}} \mathbf{x}$ is not less than zero. In addition, $\mathbf{x}^{\mathrm{T}}\sum_{s=1}^{p} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}} \mathbf{x}$ is equal to one. Therefore, we can regard the multiplications of $\mathbf{x}$ from both sides as $P(\cdot)$ and $\hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}}$ as an event in probability. In the CLAFIC method based on the discriminant rule using the entropy function, the following measurement is calculated for each class:

$$\hat{H}_k(\mathbf{x}) = -\sum_{s=1}^{p_k} \left(\mathbf{x}^{\mathrm{T}} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}} \mathbf{x}\right) \log\left(\mathbf{x}^{\mathrm{T}} \hat{\mathbf{u}}_s^k \hat{\mathbf{u}}_s^{k\mathrm{T}} \mathbf{x}\right), \tag{1}$$

where p_k is the number of basis vectors in the k-th class. We classify $\mathbf{x}$ into a class that gives the minimum value among $\hat{H}_k(\mathbf{x}) \ (k = 1,\ldots,K)$ in (1). p_k is determined by a minimum value m satisfying $\tau \le \sum_{s=1}^{m} \hat{\lambda}_s^k \Big/ \sum_{s=1}^{p} \hat{\lambda}_s^k \ (1 \le m \le p)$. To determine the optimal value of τ, for example, we can use K-fold cross-validation. The discriminant score for $\mathbf{x}_i^k$ is

$$\hat{z}(\mathbf{x}_i^k) = \hat{H}_k(\mathbf{x}_i^k) - \frac{1}{K-1} \sum_{\ell=1,\, \ell \ne k}^{K} \hat{H}_\ell(\mathbf{x}_i^k). \tag{2}$$

From (2), we calculate the average of the discriminant scores in each class as follows:

$$\hat{Z}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{z}(\mathbf{x}_i^k). \tag{3}$$

In this paper, we define $\hat{Z}_k$ in (3) as an important statistic because it represents the magnitude of separation between the k-th class and other classes.

3. Perturbation at a training sample in each class

We regard the average of the discriminant scores in the k-th class $\hat{Z}_k$ as a functional $\hat{Z}_k(\hat{F}_1, \ldots, \hat{F}_g, \ldots, \hat{F}_K)$. $\hat{F}_g$ is the empirical cumulative distribution function of the training data in the g-th class. Here we perturb $\hat{F}_g$ at a training sample in the g-th class to evaluate the influence of the training sample in each class for $\hat{Z}_k$. We denote the r-th training sample in the g-th class as $\mathbf{x}_r^g \in \mathbf{R}^p \ (r = 1, \ldots, n_g;\ g = 1, \ldots, K)$, where n_g is the number of the training samples in the g-th class. If we evaluate the influence of the r-th training sample in the g-th class for $\hat{Z}_k$ by omitting $\mathbf{x}_r^g$, we can use the sample influence function for $\hat{Z}_k$ at $\mathbf{x}_r^g$. The sample influence function for $\hat{Z}_k$ at $\mathbf{x}_r^g$ is calculated as follows:

$$\mathrm{IF}(\mathbf{x}_r^g; \hat{Z}_k) = -(n_g - 1)(\hat{Z}_k^{g(-r)} - \hat{Z}_k), \tag{4}$$

where $\hat{Z}_k^{g(-r)}$ represents $\hat{Z}_k$ when we calculate the average of the discriminant scores in the k-th class by deleting $\mathbf{x}_r^g$. In (4), the superscript $g(-r)$ means the deletion of $\mathbf{x}_r^g$, i.e., $\hat{Z}_k^{g(-r)} = 1/n_k \sum_{i=1}^{n_k} \hat{z}^{g(-r)}(\mathbf{x}_i^k)$. In this regard, $\hat{z}^{g(-r)}(\mathbf{x}_i^k)$ means $\hat{z}(\mathbf{x}_i^k)$ when we calculate the discriminant score in the i-th training sample in the k-th class for $\mathbf{x}_r^g$ deletion. In addition, $\hat{H}_g^{(-r)}(\mathbf{x}) = -\sum_{s=1}^{p_g} (\mathbf{x}^{\mathrm{T}} \hat{\mathbf{u}}_s^{g(-r)} \hat{\mathbf{u}}_s^{g(-r)\mathrm{T}} \mathbf{x}) \log(\mathbf{x}^{\mathrm{T}} \hat{\mathbf{u}}_s^{g(-r)} \hat{\mathbf{u}}_s^{g(-r)\mathrm{T}} \mathbf{x})$. When we evaluate the influence of the r-th training sample in the g-th class for $\hat{Z}_k$ by the effect of perturbation from $\hat{F}_g$ to $(1-\varepsilon)\hat{F}_g + \varepsilon \delta_{\mathbf{x}_r^g}$, we can use the empirical influence function for $\hat{Z}_k$ at $\mathbf{x}_r^g$. The empirical influence function for $\hat{Z}_k$ at $\mathbf{x}_r^g$ is calculated as follows:

$$\mathrm{IF}\left(\mathbf{x}_r^g;\hat{Z}_k\right)=\lim_{\varepsilon\to 0}\frac{\hat{Z}_k^{g\,r}-\hat{Z}_k}{\varepsilon}, \tag{5}$$

where $\hat{Z}_k^{g\,r}$ represents $\hat{Z}_k$ in the case of perturbing $\hat{F}_g$ as $(1-\varepsilon)\hat{F}_g+\varepsilon\delta_{\mathbf{x}_r^g}$. In (5), the superscript $g\,r$ means perturbation at $\mathbf{x}_r^g$. $\hat{Z}_k^{g\,r}=1/n_k\sum_{i=1}^{n_k}\hat{z}^{g\,r}\left(\mathbf{x}_i^k\right)$. $\hat{z}^{g\,r}\left(\mathbf{x}_i^k\right)$ indicates $\hat{z}\left(\mathbf{x}_i^k\right)$ when we calculate the discriminant score in the *i*-th training sample in the *k*-th class for perturbing $\hat{F}_g$ as $(1-\varepsilon)\hat{F}_g+\varepsilon\delta_{\mathbf{x}_r^g}$. $\hat{H}_g^r(\mathbf{x})=-\sum_{s=1}^{p_g}\left(\mathbf{x}^{\mathrm{T}}\hat{\mathbf{u}}_s^{g\,r}\hat{\mathbf{u}}_s^{g\,r\,\mathrm{T}}\mathbf{x}\right)\log\left(\mathbf{x}^{\mathrm{T}}\hat{\mathbf{u}}_s^{g\,r}\hat{\mathbf{u}}_s^{g\,r\,\mathrm{T}}\mathbf{x}\right)$. Then, $\hat{\mathbf{u}}_s^{g\,r}=\hat{\mathbf{u}}_s^g+\varepsilon\hat{\mathbf{u}}_s^{g(1)}+O\left(\varepsilon^2\right)$ where $\hat{\mathbf{u}}_s^{g(1)}$ is the empirical influence function of $\hat{\mathbf{u}}_s^g$ at $\mathbf{x}_r^g$. The derivation of $\hat{\mathbf{u}}_s^{g(1)}$ is given in Radhakrishnan and Kshirsagar (1981).

4. Diagnostics

In this section, we explain sensitivity analysis for the CLAFIC method based on the similarity using the entropy function.

4.1 Detection of a single influential training sample

In single-case diagnostics, we evaluate the influence of a single training sample in each class for $\hat{Z}_k$. In general, we can expect that the value of the sample influence function for $\hat{Z}_k$ at a training sample is approximately the same as the value of the empirical influence function for $\hat{Z}_k$ at the training sample. For such cases, we can evaluate the influence of the training sample in each class for $\hat{Z}_k$ $(k=1,\dots,K)$ by using (4) or (5). Concretely, we plot $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{Z}_k\right)\left(r=1,\dots,n_g\right)$ along the index *r* of the *g*-th class. Next we detect large influential training samples in each class that provide at least one large value among $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{Z}_k\right)(k=1,\dots,K)$ for each *r*. Therefore, on the basis of their training samples in each class, we can evaluate the effect of the classes for prediction accuracy. For example, we calculate all possible combinations with the detected training samples and can then perform the assessment by observing the misclassification rate on *K*-fold cross-validation while deleting each combination.

4.2 Investigation of similar influence patterns

To evaluate the influence of multiple training samples in each class for $\hat{Z}_k$, we calculate all possible combinations of the training samples in each class and assess the influence of each combination for $\hat{Z}_k$. Then, if the numbers of classes and training samples in each class are large, we perform large calculations in terms of perturbation. In this paper, we reduce the number of combinations to perturb by focusing on an important part of $\hat{Z}_k$. Here $\hat{Z}_k$ is rewritten as

$$\hat{Z}_k=\mathrm{tr}\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\left(-\sum_{s=1}^{p_k}\left(\hat{\mathbf{u}}_s^k\hat{\mathbf{u}}_s^{k\,\mathrm{T}}\right)\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^k\hat{\mathbf{u}}_s^{k\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\frac{1}{K-1}\sum_{\ell=1,\,\ell\neq k}^{K}\sum_{s=1}^{p_\ell}\left(\hat{\mathbf{u}}_s^\ell\hat{\mathbf{u}}_s^{\ell\,\mathrm{T}}\right)\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^\ell\hat{\mathbf{u}}_s^{\ell\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)\right)\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right).$$

When we perturb $\hat{Z}_k$ at $\mathbf{x}_r^g$ from the position of the empirical influence function,

$\hat{H}_g\left(\mathbf{x}_i^k\right)$ is changed into

$$-\sum_{s=1}^{p_g}\mathrm{tr}\left(\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}+\varepsilon\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}+\varepsilon\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}+O\left(\varepsilon^2\right)\right)\right.$$
$$\left.\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)+\varepsilon\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)+\varepsilon\mathrm{tr}\left(\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)+O\left(\varepsilon^2\right)\right)\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right).$$

Therefore, $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{Z}_k\right)=\lim_{\varepsilon\to 0}\left(\hat{Z}_k^{g\,r}-\hat{Z}_k\right)/\varepsilon\ (g=k)$ is

$$-\frac{1}{n_k}\mathrm{tr}\left(\sum_{i=1}^{n_k}\sum_{s=1}^{p_g}\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\right.$$
$$\left.\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}+\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\Big/\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)\right).$$

Moreover, $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{Z}_k\right)=\lim_{\varepsilon\to 0}\left(\hat{Z}_k^{g\,r}-\hat{Z}_k\right)/\varepsilon\ (g\neq k)$ becomes

$$\frac{1}{K-1}\frac{1}{n_k}\mathrm{tr}\left(\sum_{i=1}^{n_k}\sum_{s=1}^{p_g}\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\right.$$
$$\left.\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g(1)\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}+\hat{\mathbf{u}}_s^{g(1)}\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\Big/\mathrm{tr}\left(\hat{\mathbf{u}}_s^g\hat{\mathbf{u}}_s^{g\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)\right).$$

Here we denote a vector composed of the diagonal elements of

$$\frac{1}{n_k}\sum_{i=1}^{n_k}\left(-\sum_{s=1}^{p_k}\left(\hat{\mathbf{u}}_s^k\hat{\mathbf{u}}_s^{k\,\mathrm{T}}\right)\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^k\hat{\mathbf{u}}_s^{k\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)+\frac{1}{K-1}\sum_{\ell=1,\,\ell\neq k}^{K}\sum_{s=1}^{p_\ell}\left(\hat{\mathbf{u}}_s^\ell\hat{\mathbf{u}}_s^{\ell\,\mathrm{T}}\right)\log\left(\mathrm{tr}\left(\hat{\mathbf{u}}_s^\ell\hat{\mathbf{u}}_s^{\ell\,\mathrm{T}}\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}\right)\right)\right)\mathbf{x}_i^k\mathbf{x}_i^{k\,\mathrm{T}}$$

as $\hat{J}_k$. To investigate similar influence patterns of the training samples in the *r*-th training sample in the *g*-th class, we use the empirical influence function for $\hat{J}_k$, $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{J}_k\right)$. We first make a matrix that has $\mathrm{IF}\left(\mathbf{x}_r^g;\hat{J}_k\right)$ for each column and denote the matrix as [IF_g]. In addition, we estimate the asymptotic covariance matrix of $\hat{J}_k$ by using the bootstrap method. Next we calculate [IF_g]$^{\mathrm{T}}\widehat{\mathrm{acov}}\left(\hat{J}_k\right)^{-1}$[IF_g] and solve the eigenvalue problem for [IF_g]$^{\mathrm{T}}\widehat{\mathrm{acov}}\left(\hat{J}_k\right)^{-1}$[IF_g] according to Tanaka (1994). By the eigenvectors calculated from the eigenvalue problem, we search similar influence patterns of the training samples in the *g*-th class. Subsequently, we sum up the influence of the training samples in each class and reduce the number of combinations to perturb in each class. The concrete procedure of the investigation of similar influence patterns on the basis of eigenvalue problem is shown in Tanaka (1994). In particularly, the application for an extended method of linear subspace method is shown in Hayashi and Tanaka (2011).

5. Conclusions

In this paper, we focused on a linear subspace method by a similarity based on the entropy function. Through perturbation for the average of the discriminant scores in each class at the training sample, we calculated the sample and empirical influence functions for the statistics. In addition, in the CLAFIC method based on the discriminant rule by the similarity using the entropy function, we discussed the possibility of investigating similar influence patterns by focusing on the empirical influence function for the important part of the average of the discriminant scores in each class. To confirm the performance of the diagnostics suggested, we need to perform the assessment for a discriminant problem through an artificial dataset by using the CLAFIC method based on the similarity using the entropy function. In addition, we need to apply the assessment approach for a real pattern recognition problem.

Acknowledgement

This work was partly supported by Core Research of Evolutional Science & Technology (CREST) in the Japan Science and Technology Agency (Project: Alliance between Mathematics and Radiology).

References

[1] Gunal, S. and Edizkan, R. (2008). Subspace based feature selection for pattern recognition. *Information Sciences*, **178**(19), 3716-3726.

[2] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**(346), 383-393.

[3] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics: the approach based on influence functions*. John Wiley & Sons, New York.

[4] Hayashi, K. and Tanaka, Y. (2011). Sensitivity analysis for multiple similarity method and its application. In: *Proceedings of the 58th World Statistical Congress*, Dublin, 5083-5088.

[5] Maeda, E. and Murase, H. (1999). Kernel based nonlinear subspace method for pattern recognition. *Denshi Joho Tsushin Gakkai Ronbunshi*, **J82-D-II**(4), 600-612.

[6] Oja, E. (1983). *Subspace methods of pattern recognition*. Research Studies Press, England.

[7] Radhakrishnan, R. and Kshirsagar, A. M. (1981). Influence functions for certain parameters in multivariate analysis. *Communications in Statistics – Theory and Methods*, **10**(6), 515-529.

[8] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379-423, 623-656.

[9] Tanaka, Y. (1994). Recent advance in sensitivity analysis in multivariate statistical methods. *Journal of the Japanese Society of Computational Statistics*, **7**(1), 1-25.

[10] Watanabe, S. (1969). *Knowing and guessing: a quantitative study of inference and information*. John Wiley & Sons, New York.

[11] Watanabe, S. (1970). Feature compression. In: Tou, J. T. (Ed.), *Advances in information systems science vol. 3*. Plenum Press, New York-London.

[12] Watanabe, S., Lambert, P. F., Kulikowski, C. A., Buxton, J. L., and Walker, R. (1967). Evaluation and selection of variables in pattern recognition. In: Tou, J. T. (Ed.), *Computer and Information Sciences II*. Academic Press, New York.