Functional discriminant analysis for gene expression data via radial basis expansion

Araki, Yuko
Faculty of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

Imoto, Seiya
Institute of Medical Science, University of Tokyo

出版情報：COMPSTAT 2004 : Proceedings in Computational Statistics, pp.613-620, 2004. Springer
バージョン：
権利関係：

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Functional discriminant analysis for gene expression data via radial basis expansion

Y. Araki, S. Konishi
S. Imoto

MHF 2004-8

(Received March 9, 2004)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Functional discriminant analysis for gene expression data via radial basis expansion

By Yuko Araki, Sadanori Konishi

Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higasi-Ku

Fukuoka 812-8581 Japan

yuko@math.kyushu-u.ac.jp konishi@math.kyushu-u.ac.jp

And Seiya Imoto

Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku,

Tokyo 108-8639 Japan

imoto@ims.u-tokyo.ac.jp

Abstract

In this paper we introduce functional discriminant analysis which is an extension of the classical method of logistic discriminant analysis to the data where predictor variables are functions or curves. The functional discriminant analysis approach can classify curves belong to two distinct classes effectively by imposing smoothness constraint on the predictor functions and coefficient function via regularized radial basis expansion. In order to select the number of basis functions to be expanded and the value of smoothing parameter which are essential in regularization, we derive an information criterion which enables us to evaluate model estimated by regularization. The proposed method is illustrated with the example in the analysis of yeast cell cycle microarray data. It is shown that functional discriminant analysis performs well especially in the sense of prediction accuracy.

1 Introduction

Classification or discrimination technique is one of the most widely used statistical tools in various fields of natural and social sciences. In recent years, several techniques have been proposed which include penalized discriminant analysis (Hastie *et al.* (1995)) and support vector machines (Vapnik (1995)) for analyzing multivariate observations with complex structure.

Fisher's linear discriminant analysis (LDA) is a popular procedure for classification problem which is based on the Mahalanobis distances. When data are functions, however, the linear discriminant analysis cannot be applied to the discretized values directly. It is obvious that discrete observations of functions or curves are highly correlated and this statistically unfavorable situation degenerates covariance matrix and makes it impossible to take inverse of the variance-covariance matrix. Hastie *et al.* (1995) introduced penalized discriminant analysis to overcome the problems of high-dimensional correlated predictors.

They also discussed the functional version of the linear discriminant analysis, pointing out that the existence of the Mahalanobis distance in functional framework is not clear, since an inverse of the covariance operator generally cannot be represented by a kernel.

The focus in the present paper will be on the problem of classifying functions, where each observation can be interpreted as a discretized realization of a function evaluated at possibly differing time points. Our motivation arises from the analysis of yeast cell cycle gene expression data which provide inference about how gene expression levels evolve in time and how genes are dependent during a given biological process (Spellman et al. (1998), and Luan and Li (2003)). Classification of genes enables us to predict functions of unknown genes and to identify the set of co-regulated genes. In the yeast cell cycle data analysis, one wish to classify genes based on the cDNA microarray time series data.

We introduce functional discriminant analysis using Gaussian radial basis function networks with help of regularization. It is designed to construct a decision rule based on data given as a set of functions. We first transfer the vector valued observations to a set of functions. Secondly, functional discriminant analysis model is constructed by using Gaussian radial basis functions and then estimation is by regularized maximum likelihood method. In order to select smoothing parameters, we derive model selection criterion within the framework of functional data analysis by developing the generalized information criterion due to Konishi and Kitagawa (1996).

This paper is organized as follows. In Section 2 we describe the radial basis expansion smoothing technique which converts discrete raw data into underlying smooth functional form. The new method, functional discriminant analysis, is set out in Section 3 and the details of its implementation are described. Section 4 presents an application of the proposed method to yeast cell cycle gene expression data collected by Spellman et al. (1998).

2 Radial basis smoothing techniques

In the context of functional data analysis (Ramsay and Silverman (1997),(2002)), individual data should be considered to have a functional form in nature even though observed data are usually recorded discretely. In addition, those discrete raw data which are supposed to have functional form may contain observational error. Therefore, converting raw data into underlying smooth functional form requires an efficient smoothing technique.

The typical functional data analysis approach is to fit each curve individually using expansion of basis functions. Common basis functions for smoothing functional data are B-spline basis and Fourier expansions. In our model, we use Gaussian radial basis function with hyperparameter that controls the amount of overlapping among basis functions and adopts the information of the desired outputs (Ando et al. (2002)). For background about radial basis function networks, we refer to Moody and Darken (1989), Poggio and Girosi (1990), Webb (1999) and references given therein.

Suppose we have N independent observations $\{(x_i, t_i); t_i \in \mathcal{T}, i = 1,2,\cdots, N\}$, where x_i are random response variables and t_i are explanatory variables, assuming that they are drawn from the Gaussian nonlinear regression model

$$x_i = u(t_i) + \epsilon_i, \qquad i = 1, \cdots, N, \tag{1}$$

where $u(t)$ is a smooth function to be estimated, and the errors ϵ_i are independently, normally distributed with mean zero and variance σ^2. We consider the function $u(t)$ that can be expanded in the form of the radial basis function network taking the following form;

$$u(t;\boldsymbol{\omega}) = \sum_{k=1}^{m} \omega_k \phi_k(t) + \omega_0, \tag{2}$$

where $\boldsymbol{\omega} = (\omega_0, \omega_1, 2, \cdots, \omega_m)^T$ and $\phi_k(t)$ are a set of Gaussian radial basis functions with hyperparameter ν given as

$$\phi_k(t; c_k, s_k^2) = \exp\left\{-\frac{(t-c_k)^2}{2\nu s_k^2}\right\}, \qquad k = 1, 2, \cdots, m, \tag{3}$$

where c_k is a scalar determining the location of the kth basis function, s_k is the width and ν is a hyperparameter. The function $\hat{u}(t) \equiv x(t)$ which is estimated from the observed data $\{(x_i, t_i); i = 1, 2, \cdots, N\}$ is called 'functional data', and is proceeded to further analysis.

The nonlinear function $u(t)$ is estimated in two-stage procedure; position the centers and determine the dispersions first, then calculate the weights using an appropriate optimization schemes. This two stage learning is reported to solve the problem of convergence and the identification problem. Among several strategies, k-means clustering method algorithm is used to determine the centers c_k and the dispersion parameters s_k^2 of the basis functions. More precisely, observation points $\{t_1, t_2, \cdots, t_N\}$ are grouped into m clusters $\{C_1, C_2, \cdots, C_m\}$, where m is a given number of radial basis functions. Then the centers and the dispersion parameters are determined by

$$c_k = \frac{1}{n_k} \sum_{t_i \in C_k} t_i, \qquad\qquad s_k^2 = \frac{1}{n_k} \sum_{t_i \in C_k} (t_i - c_k)^2,$$

where n_k represents the number of data which belong to the cluster C_k. We define the basis function $\phi_k(t; c_k, s_k^2)$ using these estimates as $\phi_k(t)$. Hence it follows that the nonlinear regression model based on the radial basis function network can be written as

$$f(x_i|t_i; \boldsymbol{\omega}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(x_i - \boldsymbol{\omega}^T \boldsymbol{\phi}(t_i)\right)^2}{2\sigma^2}\right\}, \tag{4}$$

where $\boldsymbol{\phi}(t_i) = (1, \phi_1(t_i), \cdots, \phi_m(t_i))^T$.

In fitting data with complex structure, the maximum likelihood method does not yield satisfactory results, since it often occurs overfittng and yields unstable parameter estimates. Therefore the unknown weights and the error variances are estimated by regularization method. Moreover, in smoothing functional data, all individual data should be fitted by using the common basis functions in our model. In other words, the number of basis functions is fixed even though the amount of smoothness imposed on a set of discrete data will be differ from each other. Regularization allows us to adjust individual differences by a smoothing parameter. In addition, implementing the hyperparameter and adjusting the smoothing parameter capture the structure in the data flexibly.

The regularization method maximizes the penalized log-likelihood function

$$l_\gamma(\boldsymbol{\omega}, \sigma^2) = \sum_{i=1}^{N} \log f(x_i|t_i; \boldsymbol{\omega}, \sigma^2) - \frac{N\gamma}{2}\boldsymbol{\omega}^T D_2^T D_2 \boldsymbol{\omega}, \tag{5}$$

where $D_2^T D_2$ is the second order difference matrix and γ is called a smoothing parameter which adjusts the amount of smoothness and also avoids ill-posed problem. The maximum penalized likelihood estimates are

$$\hat{\boldsymbol{\omega}} = (\Phi^T\Phi + N\beta D_2^T D_2)^{-1}\Phi^T \boldsymbol{x}, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\{x_i - \hat{\boldsymbol{\omega}}^T \boldsymbol{\phi}(t_i)\}^2, \tag{6}$$

where $\Phi = (\boldsymbol{\phi}(t_1), \boldsymbol{\phi}(t_2), \cdots, \boldsymbol{\phi}(t_N))^T$, $\beta = \gamma\hat{\sigma}^2$ and $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)^T$. The number of basis functions m, the adjusted parameters ν and γ are determined by using a model selection criterion obtained by Ando et al. (2002). Thus the observed discrete data $\{(x_i, t_i); t_i \in \mathcal{T}, i = 1, 2, \cdots, N\}$ are smoothed by the method described above and we have a functional data given by $x(t)$;

$$\hat{u}(t) = \sum_{k=1}^{m} \hat{\omega}_k \phi_k(t) + \hat{\omega}_0 \equiv x(t), \qquad t \in \mathcal{T}. \tag{7}$$

3 Functional logistic discrimination

Suppose we have n independent observations $\{(x_\alpha(t),\ g_\alpha);\ \alpha = 1, 2, \cdots, n\}$, where $x_\alpha(t)$ are functional predictor variables and g_α are indicators of the group membership. For example, we consider two-class classification, i.e. $k = 1$ or 2, where $g_\alpha = k$ implies that it belongs to class G_k. A set of functions smoothed by the Gaussian radial basis function smoothing method are given by

$$x_\alpha(t) = \boldsymbol{w}_\alpha^T \boldsymbol{\phi}(t), \qquad \alpha = 1, 2, \cdots, n, \tag{8}$$

where $\boldsymbol{w}_\alpha$ are estimated parameter vectors and $\boldsymbol{\phi}(t)$ is a vector of Gaussian basis functions given in equation (3).

A Bayes rule of allocation is to assign $x_\alpha(t)$ to group $G_k (k = 1, 2)$ with the maximum posterior probability $\Pr(g = k|x_\alpha(t))$. We consider the log-odds of the posterior probability given in the following form;

$$\log\left\{\frac{\Pr(g = 1|x_\alpha(t))}{\Pr(g = 2|x_\alpha(t))}\right\} = \beta_a + \int_{\mathcal{T}} \beta(t) x_\alpha(t) dt. \tag{9}$$

By making use of the same Gaussian radial basis function $\boldsymbol{\phi}(t)$ as in (8), we expand the functional parameter as $\beta(t) = \beta_0 + \sum_{i=1}^{m} \beta_i \phi_i(t) = \boldsymbol{\beta}_t^T \boldsymbol{\phi}(t) (t \in \mathcal{T})$, where $\boldsymbol{\beta}_t = (\beta_0, \beta_1, \beta_2, \cdots, \beta_m)^T$. We denote the posterior probability $\Pr(g = 1|x_\alpha(t)) = \pi(x_\alpha(t))$, so that $\Pr(g = 2|x_\alpha(t)) = 1 - \pi(x_\alpha(t))$. Then the log-odds model (9) can be expressed as

$$\log\left\{\frac{\pi(x_\alpha(t))}{1 - \pi(x_\alpha(t))}\right\} = \boldsymbol{z}_\alpha^T \boldsymbol{\beta}, \tag{10}$$

where $\boldsymbol{\beta} = (\beta_a, \boldsymbol{\beta}_t^T)^T$ and Z is an $n \times (m+2)$ matrix given by

$$Z^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \Phi^T \boldsymbol{w}_1 & \Phi^T \boldsymbol{w}_2 & \cdots & \Phi^T \boldsymbol{w}_n \end{bmatrix} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_n), \tag{11}$$

with $(m+1) \times (m+1)$ matrix Φ having $\phi_{jk} = \int_{\mathcal{T}} \phi_j(t)\phi_k(t)dt, j, k = 0, 1, 2, \cdots, m$, as the (j, k)-th element.

We define the binary variable y_α coded as either 0 or 1 to indicate the group membership of a sample, where $y_\alpha = 1$ if $g_\alpha = 1$ and $y_\alpha = 0$ if $g_\alpha = 2$. The log-likelihood function is

$$l(\boldsymbol{\beta}) = \sum_{\alpha=1}^{n} \left[y_\alpha \log \pi(x_\alpha(t)) + (1 - y_\alpha) \log\{1 - \pi(x_\alpha(t))\} \right], \tag{12}$$

where $\pi(x_\alpha(t)) = \exp(\boldsymbol{z}_\alpha^T \boldsymbol{\beta})/\{1 + \exp(\boldsymbol{z}_\alpha^T \boldsymbol{\beta})\}$. We estimate the parameter vector $\boldsymbol{\beta}$ by maximizing the penalized log-likelihood function

$$l(\boldsymbol{\beta}) - \frac{n\lambda}{2} \boldsymbol{\beta}^T \boldsymbol{\beta}. \tag{13}$$

We obtain the solution $\hat{\boldsymbol{\beta}}_\lambda$ by the iterative algorithm like Newton-Raphson algorithm;

$$\boldsymbol{\beta}^{new} = \boldsymbol{\beta}^{old} - \left(\frac{\partial^2 l_\lambda(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right)^{-1} \frac{\partial l_\lambda(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}. \tag{14}$$

The crucial issue on regularization method is the choice of the optimal value of smoothing parameter λ. We obtain an information-theoretic criterion within the framework of functional data analysis. An information criterion for evaluating functional discrimination model estimated by regularization is of the form

$$\mathrm{GIC}_{\mathrm{F}} = -2 \log l(\hat{\boldsymbol{\beta}}_\lambda) + 2 \mathrm{tr} Q R^{-1}, \tag{15}$$

where Q and R are $(m+2) \times (m+2)$ matrices given by the first and second derivatives of equation (13), given by

$$Q = \frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial \{\log f(y_\alpha | x_\alpha(t); \boldsymbol{\beta}) - (\lambda/2) \boldsymbol{\beta}' \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta}} \frac{\partial \log f(y_\alpha | x_\alpha(t); \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \Bigg|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}$$

$$R = -\frac{1}{n} \sum_{\alpha=1}^{n} \frac{\partial^2 \{\log f(y_\alpha | x_\alpha(t); \boldsymbol{\beta}) - (\lambda/2) \boldsymbol{\beta}' \boldsymbol{\beta}\}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \Bigg|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}.$$

We choose the smoothing parameter λ to minimize $\mathrm{GIC}_{\mathrm{F}}$.

Schwarz (1978) proposed the Bayesian information criterion, BIC. However, the BIC covers only models estimated by the maximum likelihood method. Konishi et al. (2004) generalized BIC and derived extended BIC which enabled us to evaluate models estimated by the method of regularization. They also derived the generalized BIC for radial basis function network logistic regression model. We simply modify it to functional case, obtaining

$$BIC_{fl}(\lambda) = 2 \textstyle\sum_{\alpha=1}^{n} \left[\log \left\{ 1 + \exp\left(\boldsymbol{z}'_\alpha \hat{\boldsymbol{\beta}} \right) \right\} - y_\alpha \boldsymbol{z}'_\alpha \hat{\boldsymbol{\beta}} \right] + n\lambda \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{\beta}} - 2 \log(2\pi/n) + \log |R| - (m+1) \log \lambda.$$

4 Real data example

In this section we show the effectiveness of the proposed method through the analysis of the yeast cell cycle gene expression data collected by Spellman et al. (1998). Gene expressions for all 6,178 genes in the yeast genome were measured by cDNA microarrays over time during about two cell cycles. These data contain 77 microarrays and consist of two short time-courses (two time points) and four medium time-courses (18, 24, 17 and 14 time points). Spellman et al. (1998) identified 800 genes as cell cycle related genes based on the clustering analysis, and also grouped these genes into five classes, G1, S, G2, M, and M/G1, by considering peaks in the expression patterns.

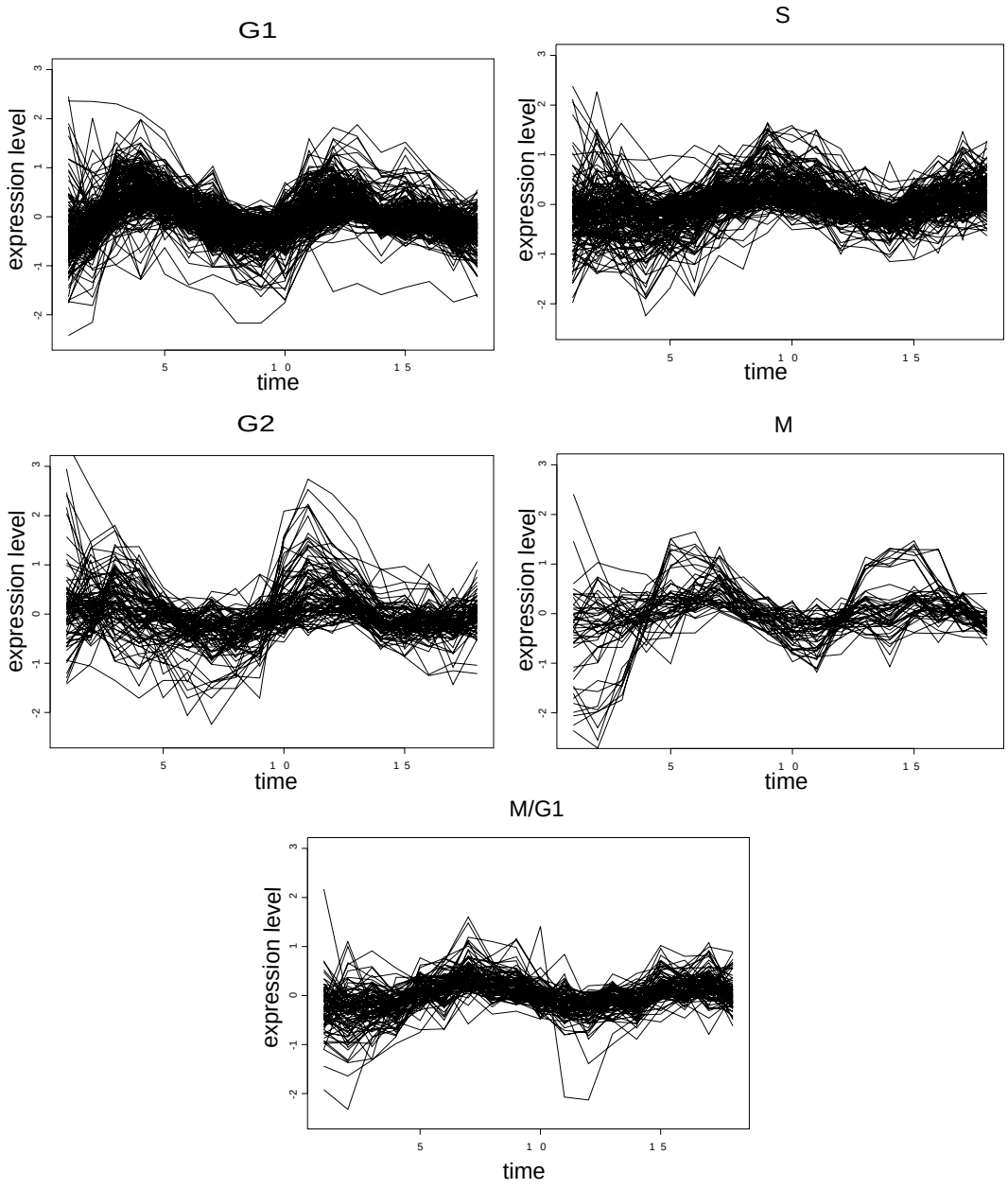


Fig. 1: Raw gene expression profiles during the yeast α factor-based synchronization experiment.

Figure 1 are plots of expression patterns of the 612 genes in the five classes, G1, S, G2, M, and M/G1 for comparison. In our analysis, we concentrated on the time-course "α factor-based synchronization experiment data" (18 time points), excluding the genes containing missing values for simplicity. That was, the expression patterns of 612 genes out of 800 cell cycle related genes were used in our analysis, and those expression data were considered as a discretized realization of 612 expression curves evaluated at 18 time points.

Note that microarray data usually contain observational noise. Therefore, the smoothing performed at first had an important role to remove the observational noise from expression data. In addition, since the gene expression pattern of each cell cycle related gene can be considered as a function of time, the proposed method is appropriate for analyzing time course gene expression data. We carried out two-class functional discrimination for all possible combinations. In order to evaluate the effectiveness of our model, the genes in each class were randomly assigned into training data and test data. That is, the model was estimated by using the training data, and the predictive accuracy of the estimated model was evaluated by the test data. We first applied the Gaussian radial basis smoothing method described in Section 2 to the time-course expression data $\{(x_{ij}, t_j,); i = 1, 2, ..., 612; j = 1, 2, ..., 18\}$, where x_{ij} is the expression value of ith gene at time t_j. As we mentioned above, the smoothing parameter and the hyperparameter are to be adjust individual smoothness differences, and they worked efficiently here in practice.

Figure 2 shows examples of two genes from G1 class with the Gaussian radial basis smoothing curves.

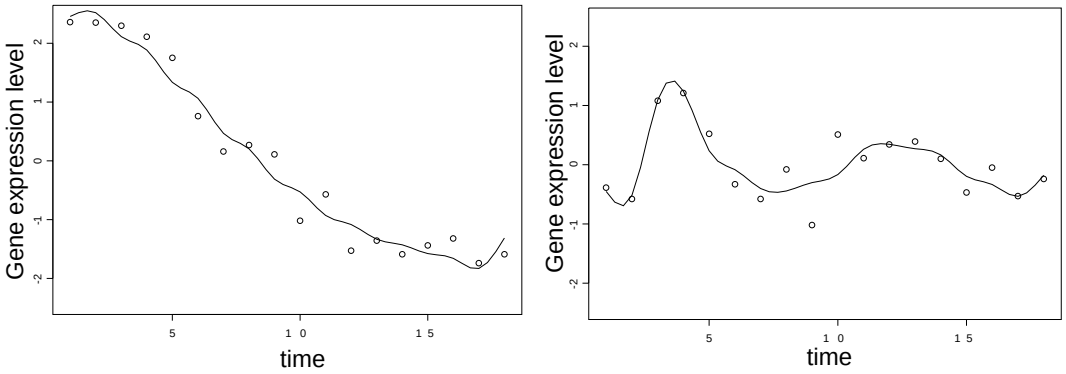


Fig. 2: Smoothed gene expression patterns of the class G1 by the Gaussian radial basis function networks with hyperparameter.

We succeeded in extracting the effective expression curves that are possibly close to the real expression patterns, even though there are various types of expression patterns in the same class. We observed that the hyperparameter allowed flexible curve fitting, while the smoothing parameter adjusted the differences of gene expression patterns. The linear discriminant analysis (LDA) and the quadratic discriminant analysis (QDA) are the most popular classical method for classification. To compared functional discriminant analysis model(FLDA) evaluated by the criterion $\mathrm{GIC_F}$ and $\mathrm{BIC_p}$ with LDA and QDA which analyzed discretized data directly, we performed two-group classification.

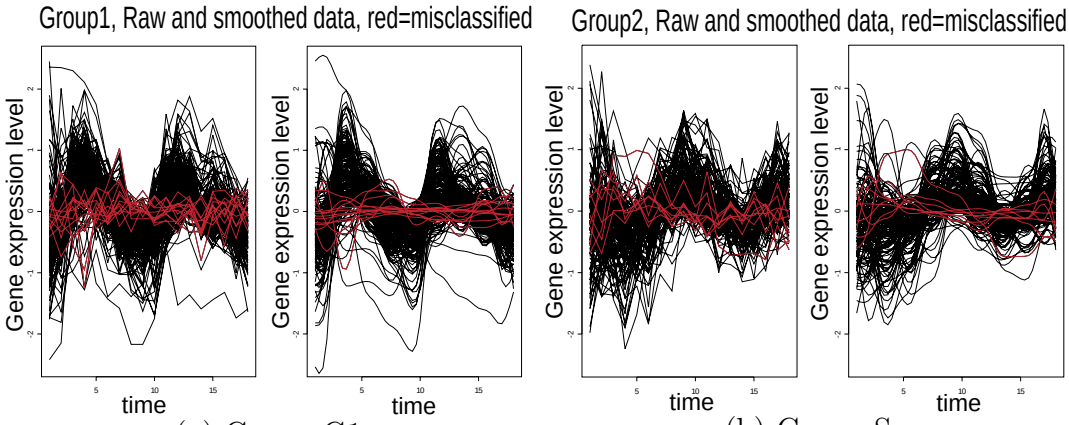
Table 1 summarizes the classification results of four different methods (LDA, QDA, FLDA with $\mathrm{GIC_F}$ and $\mathrm{BIC_p}$). For almost all combinations of the classes, the proposed

		Training	Test			Training	Test
G1 and S	LDA	0.04	0.08	S and M	LDA	0.03	0.13
	QDA	0.05	0.11		QDA	0.11	0.19
	FLDA GIC_F	0.04	0.06		FLDA GIC_F	0.08	0.1
	FLDA BIC_p	0.04	0.06		FLDA BIC_p	0	0.14
G1 and G2	LDA	0.13	0.15	S and	LDA	0.13	0.31
	QDA	0.15	0.22	M/G1	QDA	0.15	0.35
	FLDA GIC_F	0.13	0.15		FLDA GIC_F	0.12	0.28
	FLDA BIC_p	0.13	0.14		FLDA BIC_p	0.12	0.25
G1 and M	LDA	0.07	0.18	G2 and M	LDA	0.04	0.1
	QDA	0.12	0.13		QDA	0.05	0.19
	FLDA GIC_F	0.08	0.13		FLDA GIC_F	0	0.09
	FLDA BIC_p	0.08	0.13		FLDA BIC_p	0	0.1
G1 and	LDA	0.04	0.1	G2 and	LDA	0.01	0.09
M/G1	QDA	0.11	0.11	M/G1	QDA	0.03	0.12
	FLDA GIC_F	0.05	0.09		FLDA GIC_F	0.01	0.11
	FLDA BIC_p	0.05	0.1		FLDA BIC_p	0.01	0.08
S and G2	LDA	0.09	0.14	MandM/G1	LDA	0.05	0.21
	QDA	0.09	0.18		QDA	0.1	0.3
	FLDA GIC_F	0.12	0.14		FLDA GIC_F	0.16	0.24
	FLDA BIC_p	0.12	0.14		FLDA BIC_p	0.18	0.25

Table 1: Error rates of two-group discrimination with four different methods

method yielded lower test error rates. We suggest investigating genes that were classified in the opposite group with high posterior probability, since they possibly misclassified by Spellman et al. (1998).

Figure 3 plot the results of FLDA of group G1 and S. We found that most of those

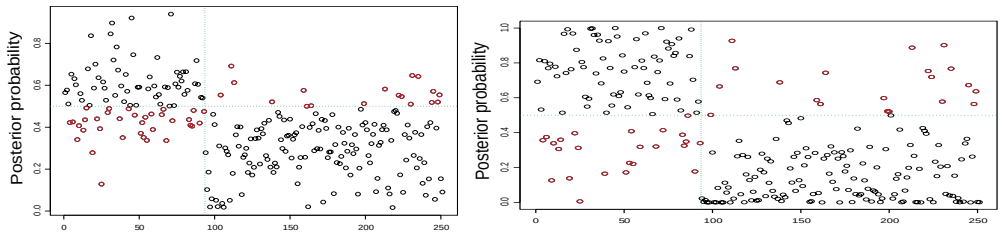


(a) Group G1 (b) Group S

Fig. 3: Group (a)G1, (b)S raw (left) and smoothed (right) gene expression data. Red lies are plots of misclassified genes.

misclassified curves have their posterior probabilities close to 0.5, which coincide with the curves of low magnitudes. Also there are one or two misclassified genes which have the posterior probability to be classified to the opposite group.

Other than the parameters described so far, we found that the convergence criterion in Newton-Raphson iteration for estimating $\hat{\beta}$ also affect the prediction accuracy. We have



(a) the convergence criterion $= 10^{-0.5}$

(b)the convergence criterion $= 10^{-2}$

Fig. 4: The posterior probabilities of genes for classification of class S and M/G1. Red points represents misclassified genes.

to determine the way to define the convergence condition in Newton-Raphson iteration in β estimation. Figure 4 compares the postserior probability for each gene when convergence criterion was altered from (a)$10^{-0.5}$ to (b)10^{-2}. In this example, the test error rate was dropped from 20.7 percent to 17.5 percent as the convergence criterion getting smaller. However, we should have some points of convergence restriction which drop the misclassification rate. The smaller value of restriction, 10^{-3}, did not drop the error rate anymore. In figure 5, we see that the strong restriction on convergence criterion seemed

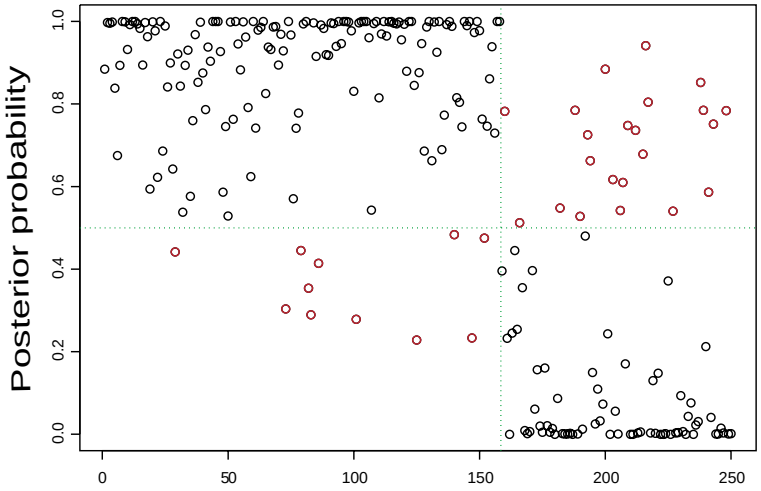


Fig. 5: The posterior probabilities when convergence criterion $= 10^{-3}$

to make most of the posterior probabilities close to 0, yet the remaining "not close to zero" genes tend to be misclassified. We need theoretical inspection about this matter.

5 Discussion

The functional discriminant analysis model proposed in this paper appeared to be a useful tool for classifying functions or curves. An advantage of our method is that one could treat the samples as a set of functions, hence the problems of the observational point difference and highly correlated data are overcome. Also the model selection criterion $\mathrm{GIC_F}$ and $\mathrm{BIC_p}$ enables us to evaluate models subjectively. Potential research would

be that extending our modeling strategy to the case of sampled surface for multi-group classification.

References

[1] Ando, T. and Konishi, S. (2002). Nonlinear regression modeling via regularized radial basis function networks. The Institute of Statistical Mathematics, Research Memorandum. **845**.

[2] Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.* **23**, 73 – 102.

[3] Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag, New York.

[4] Konishi, S. and Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika* **83**, 875 – 890.

[5] Konishi, S., Ando, T. and Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika* **91**,1,27 – 43.

[6] Luan, Y. and Li, H. (2003). Clustering of time-course gene expression data using a mixed-effects model with B-splines. *Bioinformatics* **19(4)**, 474 – 482.

[7] Moody, J. and Darken, C. J. (1989). Fast learning in networks of locally-tuned processing units. *Neural Comp.* **1**, 281 – 294.

[8] Poggio, T. and Girosi, F. (1990). Networks for approximation and learning. *Proc. IEEE* **78**, 1484 – 1487.

[9] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.

[10] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.

[11] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461 – 464.

[12] Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Bostein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell.* **9**, 3273 – 3297.

[13] Vapnik, V. N. (1995). *The nature of staitistical learning theory*. Springer-Verlag, New York.

[14] Webb, A. (1999). *Statistical Pattern Recognition*. Arnold, London.

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF

2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise

2004-2 Rui YAMAGUCHI, Eiko TSUCHIYA & Tomoyuki HIGUCHI
State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors

2004-3 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Cubic pencils and Painlevé Hamiltonians

2004-4 Atsushi KAWAGUCHI, Koji YONEMOTO & Takashi YANAGAWA
Estimating the correlation dimension from a chaotic system with dynamic noise

2004-5 Atsushi KAWAGUCHI, Kentarou KITAMURA, Koji YONEMOTO, Takashi YANAGAWA & Kiyofumi YUMOTO
Detection of auroral breakups using the correlation dimension

2004-6 Ryo IKOTA, Masayasu MIMURA & Tatsuyuki NAKAKI
A methodology for numerical simulations to a singular limit

2004-7 Ryo IKOTA & Eiji YANAGIDA
Stability of stationary interfaces of binary-tree type

2004-8 Yuko ARAKI, Sadanori KONISHI & Seiya IMOTO
Functional discriminant analysis for gene expression data via radial basis expansion