Semi-supervised logistic discrimination via graph-based regularization

Kawano, Shuichi
Graduate School of Mathematics, Kyushu University

Misumi, Toshihiro
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

出版情報：Neural processing letters. 36 (3), 2012-12. Springer
バージョン：
権利関係：The original publication is available at www.springerlink.com

MI Preprint Series

Kyushu University
The Global COE Program
Math-for-Industry Education & Research Hub

Semi-supervised logistic discrimination via graph-based regularization

Shuichi Kawano, Toshihiro Misumi & Sadanori Konishi

MI 2010-6

(Received January 27, 2010)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Semi-supervised logistic discrimination via graph-based regularization

Shuichi Kawano[1*], Toshihiro Misumi[1†] and Sadanori Konishi[2]

[1] *Graduate School of Mathematics, Kyushu University,*
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.

[2] *Faculty of Mathematics, Kyushu University,*
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.

s-kawano@math.kyushu-u.ac.jp (S. Kawano)
toshihiro.misumi@jp.astellas.com (T. Misumi)
konishi@math.kyushu-u.ac.jp (S. Konishi)

Abstract: We address the problem of constructing a nonlinear model based on both classified and unclassified data sets for classification. A semi-supervised logistic model with Gaussian basis expansions along with technique of graph-based regularization method is presented. Crucial issues in our modeling procedure are the choices of tuning parameters included in the nonlinear logistic models. In order to select these adjusted parameters, we derive model selection criteria from the viewpoints of information theory and Bayesian approach. Some numerical examples are conducted to show the effectiveness of our proposed semi-supervised modeling strategies.

Key Words and Phrases: Basis expansion, Logistic discrimination, Model selection, Regularization, Semi-supervised learning.

1 Introduction

The classification or discrimination method plays a key role in various fields of research, including engineering, artificial intelligence and life science (see, e.g., Bishop, 2006; Hastie *et al.*, 2009). In practical situations such as medical diagnosis, classifying data sets may

*Research Fellow of the Japan Society for the Promotion of Science

†Present address: Data Science Dept., Astellas Pharma Inc., 3-17-1 Hasune, Itabashi-ku, Tokyo 174-8612, Japan.

require expensive tests or tasks that must be performed by a human operator. Therefore, only small classified data sets may be available, whereas unclassified data sets can be easily obtained. In addition, for the problem of predicting protein function, we have known the functions of certain proteins through several biological experiments, whereas the functions of other proteins remain unclear because the required experimental cost and effort are too great. Under these situations, a classification procedure that combines both classified and unclassified data, called semi-supervised learning, has received considerable attention in the contemporary statistics and machine learning literature (see, e.g., Chapelle *et al.*, 2006; Liang *et al.*, 2007).

A number of studies have challenged the semi-supervised learning procedures by using techniques from various fields of research, including statistics, machine learning and graph theory; e.g., a mixture model approach (Miller and Uyer, 1997; Dean *et al.*, 2006), a logistic discriminant model approach (Amini and Gallinari, 2002; Vittaut *et al.*, 2002), a graph-based approach (Kai *et al.*, 2004; Zhou *et al.*, 2004), a support vector machine approach (Bennett and Demiriz, 1998; Vapnik, 1998), a boosting approach (Bennett *et al.*, 2002; Chen and Wang, 2007) and so on. For overviews of semi-supervised learning methods, we refer to Chapelle *et al.* (2006) and references given therein.

Semi-supervised methods based on the graph-based approach presented by earlier researchers, including Kai *et al.* (2004) and Zhou *et al.* (2004), are well known as powerful tools for extracting useful information from both classified and unclassified data sets with the help of graph-based regularization. The semi-supervised methods include some tuning parameters which should be determined by any objectively methods. In many previous works of the semi-supervised methods, however, the values of the tuning parameters seems to be given subjectively, and hence it does not always lead to provide appropriate semi-supervised models. Crucial points, therefore, are the selection of the tuning parameters included in semi-supervised models.

In this article, we propose a nonlinear semi-supervised logistic model based on Gaussian basis expansions along with the technique of graph-based regularization method. In order to select the tuning parameters in the semi-supervised logistic models objectively, we

give model selection criteria from information theoretic and Bayesian viewpoints for evaluating models estimated by the method of graph-based regularization. Several numerical examples are conducted to examine the effectiveness of our modeling strategies.

The remainder of this article is organized as follows: Section 2 describes a nonlinear logistic model using Gaussian basis functions. We also provide an estimation procedure based on a graph-based regularization, which is constructed by both classified and unclassified samples, and discuss a relationship between our proposed modeling procedure and previously proposed semi-supervised methods based on graph-based approach. Section 3 introduces model selection criteria to select tuning parameters in the logistic models from information theoretic and Bayesian perspective. Numerical experiments are illustrated to assess the performances of our proposed semi-supervised logistic discrimination in Section 4. Section 5 presents some concluding remarks.

2 Logistic discrimination for semi-supervised classification

2.1 Nonlinear logistic model via Gaussian basis expansions

Suppose we have n_1 classified observations $\{(\boldsymbol{x}_\alpha, y_\alpha); \alpha = 1, \ldots, n_1\}$ and $(n-n_1)$ unclassified observations $\{\boldsymbol{x}_\alpha; \alpha = n_1+1, \ldots, n\}$, where $\boldsymbol{x}$ denotes the p-dimensional explanatory observations and y indicates the group membership coded as 0 or 1. We first consider the problem of constructing nonlinear logistic models with Gaussian basis functions based on classified data sets, while unclassified data sets are used in estimating the parameters included in the nonlinear logistic models.

In the logistic model, we assume that

$$\Pr(Y_\alpha = 1|\boldsymbol{x}_\alpha) = \pi(\boldsymbol{x}_\alpha), \qquad \Pr(Y_\alpha = 0|\boldsymbol{x}_\alpha) = 1 - \pi(\boldsymbol{x}_\alpha), \tag{1}$$

where Y_α is regared as a random variable distributed according to the Bernoulli distribution in the form

$$f(y_\alpha|\boldsymbol{x}_\alpha; \boldsymbol{w}) = \pi(\boldsymbol{x}_\alpha)^{y_\alpha}\{1 - \pi(\boldsymbol{x}_\alpha)\}^{1-y_\alpha}. \tag{2}$$

The nonlinear logistic model further assumes that

$$\log\left\{\frac{\pi(\boldsymbol{x}_\alpha)}{1-\pi(\boldsymbol{x}_\alpha)}\right\} = w_0 + \sum_{j=1}^{m} w_j\phi_j(\boldsymbol{x}_\alpha) = \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_\alpha), \tag{3}$$

where $\boldsymbol{w} = (w_0, w_1, \ldots, w_m)^T$ is an unknown parameter vector, $\boldsymbol{\phi}(\boldsymbol{x}) = (1, \phi_1(\boldsymbol{x}), \ldots, \phi_m(\boldsymbol{x}))^T$ is a vector of basis functions and m is the number of basis functions which is selected by model selection criteia. For basis functions $\phi_j(\boldsymbol{x})$, we use Gaussian basis functions with hyperparameter given by

$$\phi_j(\boldsymbol{x}; \boldsymbol{\mu}_j, h_j^2, \nu) = \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{\mu}_j||^2}{2\nu h_j^2}\right), \quad j = 1, \ldots, m, \tag{4}$$

where $\boldsymbol{\mu}_j$ is a p-dimensional vector that determines the center of the basis function, h_j^2 is the width parameter of the basis function and ν (> 0) is hyperparameter. In classification problems, the hyperparameter ν plays a key role in adjusting the smoothness of the decision boundary (Konishi *et al.*, 2004; Ando and Konishi, 2009).

The centers $\boldsymbol{\mu}_j$ and width parameters h_j^2 included in Gaussian basis functions in Equation (4) are generally determined by using any clustering algorithm (Moody and Darken, 1989). In particular, we employ the k-means clustering algorithm (Hartigan and Wong, 1979). Using this algorithm, we assign a set of observations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ into m clusters $\{C_1, \ldots, C_m\}$ corresponding to the number of basis functions. The centers $\boldsymbol{\mu}_j$ and the width parameters h_j^2 are, respectively, determined by

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{n_j}\sum_{\boldsymbol{x}_\alpha \in C_j} \boldsymbol{x}_\alpha \quad \text{and} \quad \hat{h}_j^2 = \frac{1}{n_j}\sum_{\boldsymbol{x}_\alpha \in C_j} ||\boldsymbol{x}_\alpha - \hat{\boldsymbol{\mu}}_j||^2, \tag{5}$$

where n_j is the number of observations that belongs to the j-th cluster C_j. Replacing $\boldsymbol{\mu}_j$ and h_j^2 with $\hat{\boldsymbol{\mu}}_j$ and $\hat{h}_j^2$, respectively, we obtain a set of m basis functions given by

$$\phi_j(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_j, \hat{h}_j^2, \nu) = \exp\left(-\frac{||\boldsymbol{x}-\hat{\boldsymbol{\mu}}_j||^2}{2\nu \hat{h}_j^2}\right), \quad j = 1, \ldots, m. \tag{6}$$

The hyperparameter ν is determined by model selection criteia given in Section 3.

From the above assumptions, the log-likelihood function of nonlinear logistic models given in n_1 observations is

$$\begin{aligned}\ell(\boldsymbol{w}) &= \sum_{\alpha=1}^{n_1} \left[y_\alpha \log \pi(\boldsymbol{x}_\alpha) + (1-y_\alpha)\log\{1-\pi(\boldsymbol{x}_\alpha)\}\right] \\ &= \sum_{\alpha=1}^{n_1} \left[y_\alpha \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_\alpha) - \log\{1+\exp(\boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_\alpha))\}\right].\end{aligned} \tag{7}$$

2.2 Estimation

The maximum likelihood estimator of an unknown parameter vector $\boldsymbol{w}$ can be obtained by a maximization of the log-likelihood function (7). In this article, however, we aim to construct nonlinear logistic models by using both classified and unclassified data sets. In order to obtain the modeling procedure, we employ a graph-based regularization method in estimation procedure.

2.2.1 Graph Laplacian

One of the most important techniques in graph-based regularization methods is the use of a graph Laplacian in the regularization term. The graph Laplacian is based on a weighted adjacency matrix and a degree matrix. A weighted matrix W is an $n \times n$ matrix in the following:

$$W_{ij} = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}{2\sigma^2}\right), \quad i, j = 1, \ldots, n, \tag{8}$$

where σ^2 is a width parameter. In our study, we utilize the weighted adjacency matrix with the parameter $\sigma = \bar{h} = \sum_{i=1}^{m} \hat{h}_i/m$, where $\hat{h}_i$ is the estimated width parameter included in Gaussian basis functions in Section 2.1. We also define the degree matrix D, which is the diagonal matrix with (i, i)-th element being $D_{ii} := \sum_{j=1}^{n} W_{ij}$.

Under these notations and terms, we define the unnormalized graph Laplacian $\mathcal{L}$ as follows:

$$\mathcal{L} = D - W. \tag{9}$$

The normalized graph Laplacian L is obtained by

$$L = D^{-1/2}\mathcal{L}D^{-1/2} = I_n - D^{-1/2}WD^{-1/2}, \tag{10}$$

where I_n is an $n \times n$ identity matrix. In our proposed modeling methodologies, we use the normalized graph Laplacian. The graph Laplacian has various mathematical and statistical properties; e.g., a Laplacian operator on discrete data. For more details and theories of graph Laplacian, we refer to Chung (1997), Chapelle *et al.* (2006, Chapter 13) and von Luxburg *et al.* (2008). Henceforth, we describe the normalized graph Laplacian as the graph Laplacian for simplicity.

2.2.2 Estimation via graph Laplacian

By using the graph Laplacian L introduced by Section 2.2.1, we propose to maximize a following regularized log-likelihood function

$$\ell_\lambda(\boldsymbol{w}) = \ell(\boldsymbol{w}) - \frac{\lambda}{2n_1}\boldsymbol{w}^T\Phi^T L\Phi\boldsymbol{w}, \tag{11}$$

where $\Phi = (\boldsymbol{\phi}(\boldsymbol{x}_1), \ldots, \boldsymbol{\phi}(\boldsymbol{x}_n))^T$ and λ (> 0) is a regularization parameter. It is observed that the first term of the Equation (11) includes only classified data, while the second term consists of both classified and unclassified data sets.

The first derivative of the regularized log-likelihood function in the Equation (11) is given by

$$\frac{\partial \ell_\lambda(\boldsymbol{w})}{\partial \boldsymbol{w}} = \sum_{\alpha=1}^{n_1} \{y_\alpha - \pi(\boldsymbol{x}_\alpha)\}\, \boldsymbol{\phi}(\boldsymbol{x}_\alpha) - \frac{\lambda}{n_1}\Phi^T L\Phi\boldsymbol{w}. \tag{12}$$

However, since the likelihood equation $\partial\ell_\lambda(\boldsymbol{w})/\partial\boldsymbol{w} = \boldsymbol{0}$ is nonlinear with respect to the parameter vector $\boldsymbol{w}$, any iterative algorithms are used to estimate the parameter vector $\boldsymbol{w}$. In our modeling strategies, we employ Newton-Raphson method (see, Green and Silverman, 1994); that is, starting from an initial value, we numerically obtain a solution in the following update quantities:

$$\boldsymbol{w}^{new} = \boldsymbol{w}^{old} - \left\{\frac{\partial^2\ell_\lambda(\boldsymbol{w}^{old})}{\partial\boldsymbol{w}\partial\boldsymbol{w}^T}\right\}^{-1} \frac{\partial\ell_\lambda(\boldsymbol{w}^{old})}{\partial\boldsymbol{w}}, \tag{13}$$

where the required second derivative of the regularized log-likelihood function $\partial^2\ell_\lambda(\boldsymbol{w})/\partial\boldsymbol{w}\partial\boldsymbol{w}^T$ is given by

$$\frac{\partial^2\ell_\lambda(\boldsymbol{w})}{\partial\boldsymbol{w}\partial\boldsymbol{w}^T} = -\sum_{\alpha=1}^{n_1} \pi(\boldsymbol{x}_\alpha)\{1-\pi(\boldsymbol{x}_\alpha)\}\boldsymbol{\phi}(\boldsymbol{x}_\alpha)\boldsymbol{\phi}^T(\boldsymbol{x}_\alpha) - \frac{\lambda}{n_1}\Phi^T L\Phi. \tag{14}$$

The $(r+1)^{th}$ estimator, $\hat{\boldsymbol{w}}^{(r+1)}$, is then updated by

$$\hat{\boldsymbol{w}}^{(r+1)} = \left\{\Phi^T\Pi^{(r)}(I_{n_1} - \Pi^{(r)})\Phi + \frac{\lambda}{n_1}\Phi^T L\Phi\right\}^{-1} \Phi^T\Pi^{(r)}(I_{n_1} - \Pi^{(r)})\boldsymbol{\xi}^{(r)}, \tag{15}$$

where $\boldsymbol{\xi}^{(r)} = \Phi\boldsymbol{w}^{(r)} + \{\Pi^{(r)}(I_{n_1} - \Pi^{(r)})\}^{-1}(\boldsymbol{y} - \Pi^{(r)}\mathbf{1}_{n_1})$ and $\Pi^{(r)}$ is an $n_1 \times n_1$ diagonal matrix with $\pi(\boldsymbol{x}_\alpha)$ for the r^{th} estimator $\hat{\boldsymbol{w}}^{(r)}$ in the α^{th} diagonal element. Here, $\boldsymbol{y} = (y_1, \ldots, y_{n_1})^T$ and $\mathbf{1}_{n_1}$ is an n_1-dimensional vector, the elements of which are all 1.

We assign a future observation $\boldsymbol{z}$ into class k ($k = 1, 2$) that has the maximum posterior probability

$$\begin{aligned} \Pr(y = 1|\boldsymbol{z}) &= \hat{\pi}(\boldsymbol{z}) = \frac{\exp\{\hat{\boldsymbol{w}}^T\boldsymbol{\phi}(\boldsymbol{z})\}}{1 + \exp\{\hat{\boldsymbol{w}}^T\boldsymbol{\phi}(\boldsymbol{z})\}}, \\ \Pr(y = 0|\boldsymbol{z}) &= 1 - \hat{\pi}(\boldsymbol{z}) = \frac{1}{1 + \exp\{\hat{\boldsymbol{w}}^T\boldsymbol{\phi}(\boldsymbol{z})\}}, \end{aligned} \tag{16}$$

where $\hat{\pi}(\boldsymbol{z})$ is the estimated conditional probability and $\hat{\boldsymbol{w}}$ is a regularized maximum likelihood estimator of $\boldsymbol{w}$.

Thus, we obtain the statistical model

$$f(y_\alpha|\boldsymbol{x}_\alpha; \hat{\boldsymbol{w}}) = \hat{\pi}(\boldsymbol{x}_\alpha)^{y_\alpha}\{1 - \hat{\pi}(\boldsymbol{x}_\alpha)\}^{1-y_\alpha}. \tag{17}$$

Note that the statistical model is constructed based on the classified and unclassified observations. The statistical model estimated by maximizing the regularized log-likelihood function depends on the number of basis functions m and the values of regularization parameter λ and hyperparameter ν involved in Gaussian basis functions. It is crucial to choose the tuning parameters in our model building strategies. In order to select the values of the tuning parameters objectively, we introduce model selection criteria in Section 3 according to information-theoretic and Bayesian viewpoints.

2.3 Relationships to previous studies

Here we observe the relationships between our proposed semi-supervised method and previously proposed semi-supervised method with graph-based regularization.

One of the most useful semi-supervised procedure using graph-based approach is the learning with local and consistency method proposed by Zhou *et al.* (2004). This methodology is given as follows. For n_1 classified observations $\{(\boldsymbol{x}_\alpha, y_\alpha); \alpha = 1, \ldots, n_1\}$ and $(n - n_1)$ unclassified observations $\{\boldsymbol{x}_\alpha; \alpha = n_1 + 1, \ldots, n\}$, we redefine a class indicator vector $\boldsymbol{y}$ given in

$$y_\alpha = \begin{cases} 1 & \text{if } \boldsymbol{x}_\alpha \text{ belongs to class 1,} \\ -1 & \text{if } \boldsymbol{x}_\alpha \text{ belongs to class 2,} \\ 0 & \text{otherwise.} \end{cases} \qquad \alpha = 1, \ldots, n, \tag{18}$$

Under these notations, Zhou *et al.* (2004) proposed to minimize the following quantity with respect to the target function f;

$$\begin{aligned} Q(\boldsymbol{f}) &= \sum_{\alpha=1}^{n}(f_\alpha - y_\alpha)^2 + \gamma \sum_{i,j=1}^{n} W_{ij}\left(\frac{1}{\sqrt{D_{ii}}}f_i - \frac{1}{\sqrt{D_{jj}}}f_j\right)^2 \\ &= (\boldsymbol{f}-\boldsymbol{y})^T(\boldsymbol{f}-\boldsymbol{y}) + \gamma \boldsymbol{f}^T(I_n - D^{-\frac{1}{2}}WD^{-\frac{1}{2}})\boldsymbol{f}, \end{aligned} \tag{19}$$

where $\boldsymbol{f} = (f_1, \ldots, f_n)^T$ and γ (> 0) is a regularization parameter. Minimizing the function in the Equation (19), we obtain the solution

$$\boldsymbol{f}^* = (I_n - tD^{-\frac{1}{2}}WD^{-\frac{1}{2}})^{-1}\boldsymbol{y}, \tag{20}$$

where $t = \gamma/(1+\gamma)$. Based on the solution $\boldsymbol{f}^* = (f_1^*, \ldots, f_n^*)^T$, the classification rule is given by

$$y_\alpha = \begin{cases} 1 & \text{if } \mathrm{sign}(f_\alpha^*) \geq 0, \\ -1 & \text{if } \mathrm{sign}(f_\alpha^*) < 0 \end{cases} \tag{21}$$

for $\alpha = 1, \ldots, n$.

Unfortunately, the semi-supervised method proposed by Zhou *et al.* (2004) predicts only the class labels of the unclassified data; i.e., the procedure cannot construct a discriminant function which classifies a future observation into one of the groups. For the problem, we have constructed the classification rule in the Equation (17) to predict a future data using logistic discrimination.

Another approach to construct a discriminant function to predict a future observation is the semi-supervised method presented by Kai *et al.* (2004). Kai *et al.* (2004) considered the minimization of the quantity in the form

$$Q'(\boldsymbol{f}) = \sum_{\alpha=1}^{n_1}(f_\alpha - y_\alpha)^2 + \xi \sum_{i,j=1}^{n} W_{ij}\left(\frac{1}{\sqrt{D_{ii}}}f_i - \frac{1}{\sqrt{D_{jj}}}f_j\right)^2, \tag{22}$$

where ξ (> 0) is a regularization parameter. Note that the first term of the right-hand side in the Equation (22) sums up only classified data, while the first term in the Equation (19) does both classified and unclassified data. Kai *et al.* (2004) assumed that the function f_α $(\alpha = 1, \ldots, n_1)$ is expanded by basis functions as follows:

$$f_\alpha = w_0 + \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}_\alpha) = \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_\alpha), \tag{23}$$

where $\boldsymbol{\phi}(\boldsymbol{x}_\alpha) = (1, \phi_1(\boldsymbol{x}_\alpha), \ldots, \phi_m(\boldsymbol{x}_\alpha))^T$ is an $(m+1)$-dimensional vector of basis functions and $\boldsymbol{w} = (w_0, w_1, \ldots, w_m)^T$ is an $(m+1)$-dimensional unknown parameter. As basis functions, Kai *et al.* (2004) utilized Gaussian basis functions.

Based on the basis expansion, we can replace the function f_α in the Equation (22) into the form

$$\begin{aligned} Q'(\boldsymbol{w}) &= \sum_{\alpha=1}^{n_1} (y_\alpha - \boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_\alpha))^2 + \xi \boldsymbol{w}^T \Phi^T (I_n - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}) \Phi \boldsymbol{w} \\ &= (\boldsymbol{y} - \Phi_{n_1}\boldsymbol{w})^T (\boldsymbol{y} - \Phi_{n_1}\boldsymbol{w}) + \xi \boldsymbol{w}^T \Phi^T (I_n - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}) \Phi \boldsymbol{w}, \end{aligned} \tag{24}$$

where $\Phi_{n_1} = (\boldsymbol{\phi}(\boldsymbol{x}_1), \ldots, \boldsymbol{\phi}(\boldsymbol{x}_{n_1}))^T$. The estimator of the parameter $\boldsymbol{w}$ can be easily obtained by setting the derivative of the function $Q'(\boldsymbol{w})$ with respect to the parameter vector $\boldsymbol{w}$ to be zero.

While the semi-supervised method proposed by Kai *et al.* (2004) is a powerful tool for classification problems, this method has some disadvantages. First, Kai *et al.* (2004) employ a squared loss function in the right-hand side of Equation (22) or Equation (24). However, the squared loss function is not feasible in classification problems, since the response variable Y for classification is assumed to be discrete variables; e.g., binary or multinomial. Second, the regularization parameter included in the semi-supervised method in Kai *et al.* (2004) is subjectively given. Meanwhile, our proposed semi-supervised model has used a logistic loss function which is optimal for classification problem compared to squared loss function. We also select the values of tuning parameters including the regularization parameter objectively by using model selection criteria given in the next section.

3 Model selection criteria

This section introduces two types of model selection criteria from information-theoretic and Bayesian viewpoints. The model selection criteria derived in this paper can be regarded as tailor-made versions of model selection criteria proposed by Konishi and Kitagawa (1996) and Konishi *et al.* (2004). For details of the model selection criteria presented by these authors, we refer to Konishi and Kitagawa (2008).

3.1 Generalized information criterion

The generalized information criterion, proposed by Konishi and Kitagawa (1996), is derived for evaluating a statistical model estimated by various methods including the regularization method. According to the result of Konishi and Kitagawa (1996, p. 876), we obtain a model selection criterion for evaluating the model estimated by the maximum regularized likelihood method with graph Laplacian given by

$$\mathrm{GIC} = -2\sum_{\alpha=1}^{n_1} \log f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{w}}) + 2\mathrm{tr}\left\{Q(\hat{\boldsymbol{w}})R^{-1}(\hat{\boldsymbol{w}})\right\}. \tag{25}$$

Here $Q(\hat{\boldsymbol{w}})$ and $R(\hat{\boldsymbol{w}})$ are, respectively, an $(m+1)\times(m+1)$ matrix in the form

$$Q(\hat{\boldsymbol{w}}) = \frac{1}{n_1}\left\{\Phi^T\hat{\Lambda}^2\Phi - \frac{\lambda}{n_1}\Phi^T L\Phi\hat{\boldsymbol{w}}\mathbf{1}_{n_1}^T\hat{\Lambda}\Phi\right\}, \tag{26}$$

$$R(\hat{\boldsymbol{w}}) = \frac{1}{n_1}\Phi^T\hat{\Pi}(I_{n_1}-\hat{\Pi})\Phi + \frac{\lambda}{n_1^2}\Phi^T L\Phi, \tag{27}$$

where $\hat{\Lambda}$ and $\hat{\Pi}$ are given by

$$\hat{\Lambda} = \mathrm{diag}[y_1 - \hat{\pi}(\boldsymbol{x}_1),\ldots,y_{n_1}-\hat{\pi}(\boldsymbol{x}_{n_1})], \tag{28}$$

$$\hat{\Pi} = \mathrm{diag}[\hat{\pi}(\boldsymbol{x}_1),\ldots,\hat{\pi}(\boldsymbol{x}_{n_1})]. \tag{29}$$

3.2 Generalized Bayesian information criterion

Consider the prior distribution for the parameter vector $\boldsymbol{w}$ given by

$$\pi(\boldsymbol{w}|\lambda) = (2\pi)^{-\{(m+1)-k\}/2}\left(\frac{\lambda}{n_1}\right)^{\{(m+1)-k\}/2}|\Phi^T L\Phi|_+^{\frac{1}{2}}\exp\left(-\frac{\lambda}{2n_1}\boldsymbol{w}^T\Phi^T L\Phi\boldsymbol{w}\right), \tag{30}$$

where $|\Phi^T L\Phi|_+$ is the product of the positive eigenvalues of $\Phi^T L\Phi$ with rank k. A mode of posterior distribution based on this prior distribution is equivalent to the maximization of the regularized likelihood function in the Equation (11).

We then derive a following model selection criterion using the result of Konishi *et al.* (2004, p. 30), which extended Schwarz's BIC (Schwarz, 1978) for evaluating the model estimated by the regularization method.

$$\begin{aligned}\mathrm{GBIC} = -2\sum_{\alpha=1}^{n_1}\log f(y_\alpha|\boldsymbol{x}_\alpha;\hat{\boldsymbol{w}}) + \frac{\lambda}{n_1}\boldsymbol{w}^T\Phi^T L\Phi\boldsymbol{w} + \log|R(\hat{\boldsymbol{w}})| \\ - \log|\Phi^T L\Phi|_+ - k\,\log\lambda - (m+1-k)\,\log(2\pi) + (m+1+k)\log n_1,\end{aligned} \tag{31}$$

where $R(\hat{\boldsymbol{w}})$ is an $(m+1) \times (m+1)$ matrix given in the Equation (27). We choose adjusted parameters including the number of basis functions m and values of regularization parameter λ and hyperparameter ν from the minimizer of the model selection criteria.

4 Numerical examples

In this section, our proposed semi-supervised logistic models with graph-based regularization are applied to several benchmark data sets.

4.1 Toy example

We confirm that models constructed by our semi-supervised procedure contains information from both classified and unclassified data by the use of a toy example. We utilized a two moon data set shown in Figure 1, where the data set $\{(x_\alpha, y_\alpha); \alpha = 1, \ldots, 110\}$ was generated from

$$
\begin{aligned}
&u_\alpha \sim U(0, 2\pi), \quad \varepsilon_{1\alpha}, \varepsilon_{2\alpha} \sim U(-0.1, 0.1), \\
&y_\alpha = 1 : z_\alpha = \cos(u_\alpha) + \varepsilon_{2\alpha},\ y_\alpha = z_\alpha \mathbf{1}_{[z_\alpha \leq 0.2]} - z_\alpha \mathbf{1}_{[z_\alpha > 0.2]},\ x_\alpha = \sin(u_\alpha) + 0.7 + \varepsilon_{1\alpha}, \\
&y_\alpha = 0 : z_\alpha = \cos(u_\alpha) + \varepsilon_{2\alpha},\ y_\alpha = z_\alpha \mathbf{1}_{[z_\alpha \geq -0.2]} - z_\alpha \mathbf{1}_{[z_\alpha < -0.2]},\ x_\alpha = \sin(u_\alpha) + \varepsilon_{1\alpha}, \quad (32)
\end{aligned}
$$

where $\mathbf{1}_{[x \geq 0]}$ denotes a indicator function. In order to perform semi-supervised inference, we considered the situation such that the data set consists of 110 examples with only 1 classified data for each class (triangle and quadrangle). For details of the data set, we refer to Zhou *et al.* (2004).

A classification boundary in the right-hand panel of Figure 1 is generated from a supervised method presented by Ando and Konishi (2009). The decision boundary fails to find the underlying data sampling structure, since the supervised logistic model is estimated by using only two classified data; that is, the supervised logistic model cannot use unclassified data sets in constructing the models. Meanwhile, the left-hand panel of Figure 1 shows the decision boundary of our proposed logistic model estimated by semi-supervised inference, where some tuning parameters included in our models are fixed as $m = 15$, $\lambda = 10^{0.3}$ and $\nu = 10$. As shown in this panel, our semi-supervised modeling

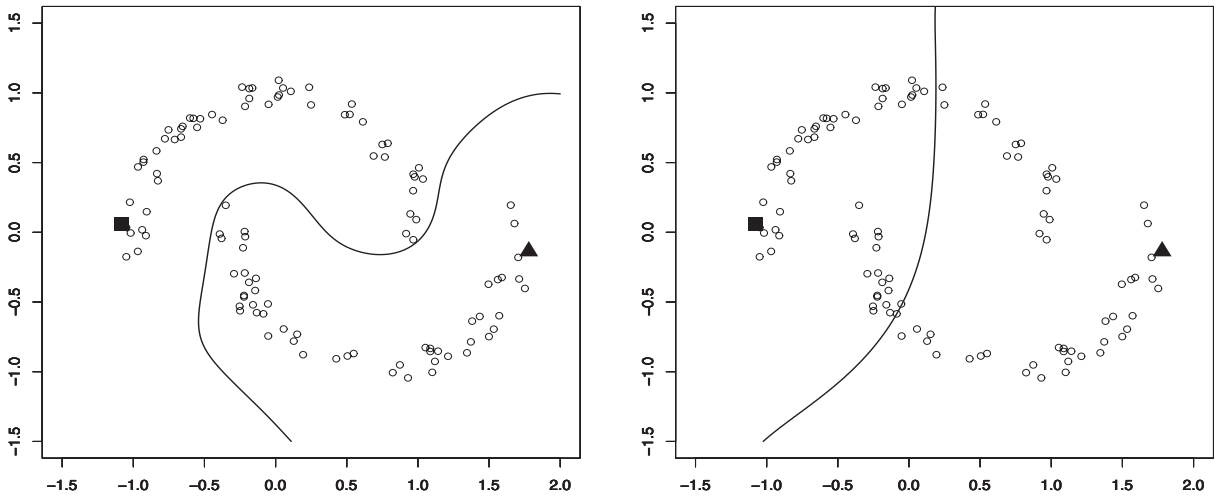

Figure 1: Two moon data set. 110 examples with only 1 labeled data for each class (triangle and quadrangle). The left panel shows the decision boundary for our proposed semi-supervised procedure with graph-based regularization. The right panel indicates the decision boundary generated from supervised logistic discrimination by Ando and Konishi (2009).

procedure correctly captures the underlying data structure thanks to the regularization term with the graph Laplacian which is based on both classified and unclassified data.

4.2 Benchmark data analysis

We examine the performances of our modeling procedure using g10 data (Chapelle and Zien, 2005) and spam data (Hastie *et al.,* 2009). The g10 data set consists of two classes with 10 predictors, and we prepared 250 sets of training data for each class and 300 sets of test data. The spam data set, which consists of 500 sets of training data and 500 sets of test data, represents the binary classification with 53 predictors. In order to implement a semi-supervised learning, the training data sets were randomly divided into two halves with classified data sets and unclassified data sets, where classified data sets were assigned as 5%, 10%, 20%, 30%, 40% and 50% of training data sets, respectively.

Our nonlinear semi-supervised logistic models estimated by graph-based regularization (SSLDA: Semi-Supervised Logistic Discriminant Analysis) were fitted to the data sets. We chose the number of basis functions m, the values of regularization parameter λ and hyperparameter ν that minimize either the GIC or the GBIC given in Section 3. The

Table 1: Comparisons of prediction error rates with different percentages of classified data sets for g10 data set. SSLDA (GBIC) or SLDA (GBIC) denotes the SSLDA or the SLDA evaluated by the GBIC, while SSLDA (GIC) or SLDA (GIC) denotes that evaluated by the GIC. Figures in parentheses for LLGC and ILLGC indicate the values of tuning parameters.

Method ╲ %	5	10	20	30	40	50
SSLDA (GBIC)	34.2	24.4	12.8	11.4	8.40	5.83
SSLDA (GIC)	27.3	25.7	13.8	13.0	9.30	7.03
LLGC (0.99)	50.0	48.1	50.8	50.6	51.1	49.1
LLGC (0.7)	47.2	41.3	38.7	41.0	33.0	25.8
LLGC (0.5)	45.1	39.6	33.9	34.2	27.6	19.2
LLGC (0.3)	43.0	37.7	30.0	28.6	24.9	16.6
LLGC (0.1)	40.8	36.0	27.2	25.6	23.1	14.8
ILLGC (100)	50.0	50.0	50.8	50.6	50.7	47.1
ILLGC (10)	50.0	45.7	45.4	45.2	35.1	25.6
ILLGC (1)	45.8	36.9	22.6	16.9	15.8	10.3
ILLGC (0.1)	30.0	26.2	14.1	11.5	11.2	8.00
ILLGC (0.01)	26.7	25.1	14.2	12.2	11.3	7.90
SVM	38.9	25.2	15.2	13.2	13.3	12.5
K-NN	35.5	31.9	27.8	25.1	23.3	23.2
SLDA (GBIC)	43.0	36.4	13.7	9.93	9.50	7.20
SLDA (GIC)	42.7	35.0	15.8	11.9	10.3	7.46

Table 2: Comparisons of prediction error rates with different percentages of classified data sets for spam data set. SSLDA (GBIC) or SLDA (GBIC) denotes the SSLDA or the SLDA evaluated by the GBIC, while SSLDA (GIC) or SLDA (GIC) denotes that evaluated by the GIC. Figures in parentheses for LLGC and ILLGC indicate the values of tuning parameters.

Method ╲ %	5	10	20	30	40	50
SSLDA (GBIC)	36.6	36.3	33.3	31.1	31.6	29.7
SSLDA (GIC)	35.6	33.3	32.8	30.6	31.2	29.3
LLGC (0.99)	36.5	36.1	37.4	36.6	37.2	37.2
LLGC (0.7)	35.6	33.4	35.1	34.5	34.4	34.3
LLGC (0.5)	35.5	33.5	34.8	33.9	34.2	34.0
LLGC (0.3)	35.0	33.6	34.4	33.4	33.8	33.5
LLGC (0.1)	34.6	33.6	34.0	33.1	33.7	33.5
ILLGC (100)	37.9	37.6	38.3	37.0	37.3	38.3
ILLGC (10)	37.0	36.2	36.9	35.7	35.8	35.8
ILLGC (1)	35.8	34.4	34.2	32.1	33.2	31.3
ILLGC (0.1)	34.5	32.3	32.4	30.4	31.6	29.5
ILLGC (0.01)	34.9	32.8	33.3	30.4	31.5	29.5
SVM	41.6	39.6	39.6	37.4	37.5	36.4
K-NN	36.8	34.3	32.3	32.0	32.7	29.2
SLDA (GBIC)	37.9	36.1	33.3	31.5	32.1	30.0
SLDA (GIC)	37.2	35.6	33.4	32.3	32.0	29.6

results for proposed modeling procedures are denoted as SSLDA(GBIC) and SSLDA(GIC) in Table 1 and Table 2.

We compared the performances of our proposed method with graph-based regularization with those of various procedures. As for other types of semi-supervised learning, semi-supervised methods proposed by Kai *et al.* (2004) (ILLGC: Inductive Learning with Local and Global Consistency) and Zhou *et al.* (2004) (LLGC: Learning with Local and Global Consistency) given in Section 2.3 were used. Since the LLGC and the ILLGC, respectively, include a tuning parameter, we set the values of the parameter into t =0.99, 0.7, 0.5, 0.3, 0.1 for the LLGC and ξ =100, 10, 1, 0.1, 0.01 for the ILLGC, respectively. Here t and ξ are defined by the Equation (20) and the Equation (22), respectively. We also employed a nonlinear logistic discrimination (SLDA: Supervised Logistic Discriminant Analysis), which is introduced by Ando and Konishi (2009), support vector machine (SVM) and k-nearest neighborhood method (K-NN). The tuning parameters in the SVM were optimized by the 5-fold cross validation, while the parameters in the K-NN were optimized by the leave one out cross validation. Note that these supervised methods are estimated by using only classified data sets. We repeated 10 times for random splits of data sets.

A summary of the prediction error rates for the g10 data set is show in Table 1, while Table 2 shows that for the spam data set. We observe that our modeling procedures using the GBIC or the GIC give relatively lower prediction errors than other semi-supervised methods and supervised methods.

5 Concluding remarks

We presented a nonlinear semi-supervised logistic model based on Gaussian basis functions along with the technique of graph-based regularization. In order to select the values of some tuning parameters, we introduced model selection criteria from information-theoretic and Bayesian approaches. Our proposed methodologies estimated by graph Laplacian are easily applied to analyze complex or high-dimensional data which include both classified and unclassified data sets. Some numerical examples illustrated that our modeling

strategies yield lower prediction error rates than previously developed methods.

In practical situations, a multi-class classification problem has attracted a great deal of attention in the various fields including statistics and machine learning (see, e.g., Lee *et al.*, 2004; Zhu *et al.*, 2009). It is easy to extent the logistic discrimination for two class classification into the discriminant model for multi-class classification. The further research remains to be done in constructing a multi-class semi-supervised discrimination with Gaussian bases with the help of graph-based regularization.

References

Amini, M-R. and Gallinari, P. (2002). Semi-supervised logistic regression. *Proceedings of the 15th European Conference on Artificial Intelligence*, 390–394.

Ando, T. and Konishi, S. (2009). Nonlinear logistic discrimination via regularized radial basis functions for classifying high-dimensional data. *Annals of the Institute of Statistical Mathematics*, **61**, 331–353.

Bennett, K. P. and Demiriz, A. (1998). Semi-supervised support vector machines. *Advances in Neural Information Processing Systems*, **11**, 368–374.

Bennett, K. P., Demiriz, A. and Maclin, R. (2002). Exploiting unlabeled data in ensemble methods. *Proceedings of ACM International Conference on Knowledge Discovery and Data Mining*, 289–296.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Chapelle, O., Schölkopf, B. and Zien, A. (2006). *Semi-Supervised Learning*. Cambridge, MA: MIT Press.

Chapelle, O. and Zien, A. (2005). Semi-supervised classification by low density separation. *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics*, 57–64.

Chen, K. and Wang, S. (2007). Regularized boost for semi-supervised learning. *Advances in Neural Information Processing Systems*, **20**, 281–288.

Chung, F. R. K. (1997). *Spectral Graph Theory*. Number 92 in Regional Conference Series in Mathematics. American Mathematical Society, Province, RI.

Dean, N., Murphy, T. B. and Downey, G. (2006). Using unlabelled data to update classification rules with applications in food authenticity studies. *Journal of the Royal Statistical Society C*, **55**, 1–14.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman & Hall.

Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, **28**, 100–108.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. 2nd ed. New York: Springer.

Kai, Y., Tresp, V. and Zhou, D. (2004). Semi-supervised induction with basis functions. Technical Report. Department of Empirical Inference, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany.

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

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

Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. New York: Springer.

Lee, Y., Lin, Y. and Wahba, G. (2004). Multicategory support vector machines, theory, and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association*, **99**, 67–81.

Liang, F., Mukherjee, S. and West, M. (2007). The use of unlabeled data in predictive modeling. *Statistical Science*, **22**, 189–205.

Miller, D. and Uyar, H. S. (1997). A mixture of experts classifier with learning based on both labelled and unlabelled data. *Advances in Neural Information Processing Systems*, **9**, 571–577.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–

464.

Vapnik, V. (1998). *Statistical Learning Theory*. New York: Wiley.

Vittaut, J-N., Amini, M-R. and Gallinari, P. (2002). Learning classification with both labeled and unlabeled data. *Proceedings of the 13th European Conference on Machine Learning*, 468–479.

von Luxburg, U., Belkin, M. and Bousquet, O. (2008). Consistency of spectral clustering. *Annals of Statistics*, **36**, 555–586.

Zhou, D., Bousquet, O., Lal, T. N., Weston, J. and Schölkopf, B. (2004). Learning with local and global consistency. *Advances in Neural Information Processing Systems*, **16**, 321–328.

Zhu, J., Zou, H, Rosset, S. and Hastie, T. (2009). Multi-class adaboost. *Statistics and Its Interface*, **2**, 349–360.

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine

MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization