Supervised image classification in Markov random field models with Jeffreys divergence

Nishii, Ryuei
Faculty of Mathematics, Kyushu University

Eguchi, Shinto
Institute of Statistical Mathematics

出版情報：MHF Preprint Series. 2004-17, 2004-06-03. 九州大学大学院数理学研究院
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

Supervised image classification in Markov random field models with Jeffreys divergence

R. Nishii & S. Eguchi

MHF 2004-17

(Received June 3, 2004)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Supervised image classification in Markov random field models with Jeffreys divergence

R. NISHII
Graduate School of Mathematics, Kyushu University
Hakozaki, Higashiku, Fukuoka 812-8581, Japan
e-mail: nishii@math.kyushu-u.ac.jp

and S. EGUCHI
Institute of Statistical Mathematics, Minami-azabu, Minatoku
Tokyo 106-8569, Japan; e-mail: eguchi@ism.ac.jp

(*Received*)

Abstract. We consider supervised image classification based on Markov Random Fields (MRFs), which are derived by the Jeffreys divergence between class-conditional probability densities. It is shown that the MRF gives an extension of Switzer's smoothing method for classification. Further, the exact error rate due to the MRF is obtained in a simple setup, and several properties of the classification are derived. Our procedure is applied to two multispectral images, and shows a good performance.

1. Introduction

Image classification is a problem of dividing an observed image or multispectral images into several homogeneous regions by labeling individual pixels based on feature information and spatial information. The problem is important and fundamental in geo-spatial data analysis because detection of land-cover categories of interest is a start point of any analysis. For this purpose, Markov random fields (MRFs) successfully model the spatial distribution of the categories. Li (2000) gave an excellent review on image analysis based on MRFs.

In image classification, we usually assume that feature vectors given category labels are independently distributed, and the labels follow an MRF. Thus the joint distribution is obtained by the product of class-conditional distributions and the joint distribution of the field. Then, the category labels are estimated by *maximum a posteriori* (MAP) principle.

The estimates of the parameters specifying class-conditional distributions can be easily obtained thanks to the conditional independence. The most difficult issue, however, is the estimation of parameters specifying the random fields since the joint distribution can not be expressed in a closed form. For this purpose, computer-intensive methods like MCMC are available, but the implementation often meets difficult aspects in the light of computational feasibility. See Mclachlan (1992), Jain et al. (2000), Eguchi and Copas (2002) for classification, Cressie (1993), Chiles (1999) for spatial statistics, and Marroquin et al. (2001), Wilson and Li (2003) for MRF-based classification/segmentation.

The Jeffreys divergence (Jeffreys 1946) offers inherited distance between class-conditional densities. In this paper, we incorporate the divergence into MRF modeling. This is an extension of the approach based on the squared Mahalanobis distance (Nishii 2003). The field based on the divergence (*the divergence model*) is specified by a unique parameter, and has a simple structure. Further, the classification rule based on the model throws a new light on Switzer's smoothing method (Switzer 1980). The divergence model is applied to benchmark data for classification, and we found it is efficient in comparison with the Ising model.

Section 2 reviews the basic distributional assumptions in image classification, and we treat MRFs determined by the divergence between the class-conditional density functions. Then, the relation between the divergence model and Switzer's smoothing method is discussed in Section 3. The error rates are exactly obtained in a simple setup. This shows that the use of spatial information always improves noncontextual classification even if the estimate of the spatial dependency parameter is far from the true value. The estimation procedure of the parameters is given at Section 4. The proposed methods are applied to benchmark data and simulated data in Section 5. Then, the divergence model shows better performance than the Ising model does. Finally, we conclude the paper at Section 6.

2. MRFs based on Jeffreys divergence

The objective of this section is to consider MRFs based on divergence (*divergence models*) for the contextual image classification.

2.1. Basic distributional assumptions on spatial data

Let $\mathcal{D} = \{1, ..., n\}$ be a training area consisting of n pixels, $\mathcal{G} = \{1, ..., G\}$ be a set of category labels, $\boldsymbol{X}_i$ be a random vector of m-dimensional feature variables at a pixel i in $\mathcal{D}$, and Y_i be a categorical variable taking values over the label set $\mathcal{G}$. Now, suppose that a set of training data $\{(\boldsymbol{x}_i, y_i) \mid i \in \mathcal{D}\}$ is available. Put $\boldsymbol{X} = (\boldsymbol{X}_1^T, ..., \boldsymbol{X}_n^T)^T$ and $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$. The joint distribution $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})$ of the feature vector $\boldsymbol{X}$ and the label vector $\boldsymbol{Y}$ can then be decomposed as $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}) = p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot p_{\boldsymbol{Y}}(\boldsymbol{y})$, where $p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y})$ is a conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$, and $p_{\boldsymbol{Y}}(\boldsymbol{y})$ is a distribution of the field. The fundamental assumption is the conditional independence of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$, that is

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) = \prod_{i \in \mathcal{D}} f(\boldsymbol{x}_i, \boldsymbol{\theta}(y_i)). \tag{1}$$

The function $f(\boldsymbol{x}, \boldsymbol{\theta})$ is called a *class-conditional probability density function* over the m-dimensional real space with the parameter vector $\boldsymbol{\theta}$.

Next, the label vector is assumed to follow a pairwise-dependent MRF with a neighborhood system $\{\mathcal{N}_i \mid i \in \mathcal{D}\}$. Let $J(g, h)$ be a quasi-distance between two categories C_g and C_h, and $\boldsymbol{Y}_{-i}$ be a vector of all labels except Y_i. Then, we assume that the conditional probability of $Y_i = g$ given $\boldsymbol{Y}_{-i} = \boldsymbol{y}_{-i}$ is specified by its neighborhood $\mathcal{N}_i$ as

$$\Pr\{Y_i = g \mid \boldsymbol{Y}_{-i} = \boldsymbol{y}_{-i}\} = \frac{\exp\{-\beta \Delta_i(g)\}}{\sum_{g' \in \mathcal{G}} \exp\{-\beta \Delta_i(g')\}} \quad \text{for} \quad g \in \mathcal{G}, \quad \text{say} \quad p_i(g \mid \boldsymbol{y}_{-i}) \tag{2}$$

where β is a non-negative constant called *a spatial dependency parameter*, and

$$\Delta_i(g) = \sum_{h \in \mathcal{G}} r_i(h) J(h, g), \quad r_i(g) \quad = \quad |\{j \in \mathcal{N}_i \mid y_j = g\}| / |\mathcal{N}_i|. \tag{3}$$

Note that $r_i(g)$ is a relative frequency of neighboring pixels labeled by g in $\mathcal{N}_i$. The Hammersley-Clifford theorem assures that the conditional distributions (2) specify the joint distribution of the label vector $\boldsymbol{Y}$ under the mild condition.

The label vector of test data will be estimated by the MAP estimate. Simulated annealing due to Geman and Geman (1984) and the iterative conditional modes (ICM) method due to Besag (1986) are the major estimation approaches.

2.2. The Ising model and Jeffreys divergence

For the quasi-distance $J(g,h)$, the simplest one would be 0-1 distance defined by $J_0(g,h) := 0$ if $g = h$, $:= 1$ if $g \neq h$. The spatial model with the distance $J_0(g,h)$ is the Ising model, and it is one of the pillars of statistical mechanics. The model, however, is not always suitable in the case having more than two categories. For example consider a geographical map with three categories of "conifer", "broad leaf" and "water". The model based on the 0-1 distance implies that $J_0(\text{“water”}, \text{“conifer”}) = J_0(\text{“broad leaf”}, \text{“conifer”})$. The equality does not fit to geography. Then, Nishii (2003) proposed to take the quasi-distance by the squared Mahalanobis distance. His approach will be extended into more general setting.

In this paper, we propose to use the Jeffreys divergence (Jeffreys 1946) of the class-conditional densities as a quasi-distance between the categories. The divergence between two probability densities is defined as follows:

$$J(g,h) = \int \{f(\boldsymbol{x}, \boldsymbol{\theta}(g)) - f(\boldsymbol{x}, \boldsymbol{\theta}(h))\} \log \frac{f(\boldsymbol{x}, \boldsymbol{\theta}(g))}{f(\boldsymbol{x}, \boldsymbol{\theta}(h))} d\boldsymbol{x} \geq 0. \tag{4}$$

This is the symmetrized Kullback-Leibler divergence. It is well-known that the equality holds if and only if $g = h$. See Taneja (1995) for further properties.

As an example, consider Gaussian distributions. Assume that the class-conditional distribution is given by a Gaussian distribution with mean vector $\boldsymbol{\mu}(g)$ and common covariance matrix Σ (*homoscedastic*) denoted by $N_m(\boldsymbol{\mu}(g), \Sigma)$ for g in $\mathcal{G}$. Then, the class-conditional density is given by

$$\psi(\boldsymbol{x}; \boldsymbol{\mu}(g), \Sigma) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\{-D(\boldsymbol{x}, \boldsymbol{\mu}(g); \Sigma)\} \tag{5}$$

where $D(\boldsymbol{s}, \boldsymbol{t}; \Sigma)$ stands for a half of the squared Mahalanobis distance:

$$D(\boldsymbol{s}, \boldsymbol{t}; \Sigma) = (\boldsymbol{s} - \boldsymbol{t})^T \Sigma^{-1} (\boldsymbol{s} - \boldsymbol{t})/2. \tag{6}$$

In the homoscedastic case, the Jeffreys divergence is reduced to the squared Mahalanobis distance $2D(\boldsymbol{\mu}(g), \boldsymbol{\mu}(h); \Sigma)$. In the heteroscedastic case, the divergence between two Gaussian distributions $N_m(\boldsymbol{\mu}(g), \Sigma(g))$ and $N_m(\boldsymbol{\mu}(h), \Sigma(h))$ is given by the formula (21) in Appendix.

3. Gaussian MRFs and error estimates

Focus our attention to a fixed pixel and its neighborhood, say 0 and $\mathcal{N}_0$. We discuss the classification problem of the center pixel 0 when neighboring labels y_j in $\mathcal{N}_0$ are given.

3.1. A relation between the divergence model and Switzer's model

Let us consider the divergence model in Gaussian MRFs (GMRFs), where feature vectors follow homoscedastic Gaussian distribution $N_m(\boldsymbol{\mu}(g), \Sigma)$. It will be shown that the divergence model in this case is a natural extension of Switzer's model.

Let $\beta_*(\geq 0)$ be a pre-supposed value of the unknown parameter β in the formula (2). We will estimate the label y_0 of the fixed pixel 0 by the ICM algorithm. In this case, the estimate is derived by maximizing the product of the densities $\psi(\boldsymbol{x}_0; \boldsymbol{\mu}(g), \Sigma)$ and the conditional probability $p_0(g \mid \boldsymbol{y}_{-i})$ defined by (5) and (2) respectively. Hence the estimate of y_0 due to the divergence model is obtained by

$$\widehat{Y}_{\text{Divergence}} = \underset{g \in \mathcal{G}}{\arg\min} \Big[D(\boldsymbol{x}_0, \boldsymbol{\mu}(g); \Sigma) + \frac{2\beta_*}{|\mathcal{N}_0|} \sum_{j \in \mathcal{N}_0} D(\boldsymbol{\mu}(y_j), \boldsymbol{\mu}(g); \Sigma) \Big], \tag{7}$$

where $D(\boldsymbol{s}, \boldsymbol{t}; \Sigma)$ is defined by (6).

Switzer's model (Switzer 1980) is based on the local continuity assumption, that is "a center pixel and its neighbors belong to the same category". Then, the center is classified by the majority vote of likelihoods by maximizing $\log \psi(\boldsymbol{x}_0; \boldsymbol{\mu}(g), \Sigma) + \sum_{j \in \mathcal{N}_0} \log \psi(\boldsymbol{x}_j; \boldsymbol{\mu}(g), \Sigma)$ w.r.t. the label g. Here, Switzer's model can be slightly extended by changing weights. The target function is switched to $\log \psi(\boldsymbol{x}_0; \boldsymbol{\mu}(g), \Sigma)$ $+2\beta_* \sum_{j \in \mathcal{N}_0} \log \psi(\boldsymbol{x}_j; \boldsymbol{\mu}(g), \Sigma)/|\mathcal{N}_0|$. This is equivalent to

$$\widehat{Y}_{\text{Switzer}} = \underset{g \in \mathcal{G}}{\arg\min} \Big[D(\boldsymbol{x}_0, \boldsymbol{\mu}(g); \Sigma) + \frac{2\beta_*}{|\mathcal{N}_0|} \sum_{j \in \mathcal{N}_0} D(\boldsymbol{x}_j, \boldsymbol{\mu}(g); \Sigma) \Big]. \tag{8}$$

The difference between two formulas (7) and (8) is found in the terms $\boldsymbol{\mu}(y_j)$ and $\boldsymbol{x}_j$ only. Note that $\boldsymbol{x}_j$ itself is a primitive estimate of $\boldsymbol{\mu}(y_j)$. Hence, the classification method based on the divergence model can be regarded as a natural extension of Switzer's method.

It is known that Switzer's method gives fairly good classification results in many practical situations. The local continuity assumption of the categories, however, cannot be applied to the whole image, and the estimation procedure of the parameters is not well-established. Thus, the divergence model can be seen as an extension of Switzer's method into the established MRF-based framework.

3.2. Error rates due to the divergence model and Switzer's model

We will derive the exact error rate due to the divergence model and Switzer's model in the previous local region with two categories $\mathcal{G} = \{1, 2\}$. In the two-category case, a positive quasi-distance is $J(1, 2)$ alone. Hence, by replacing $\beta J(1, 2)$ for β, the MRF based on Jeffreys divergence is reduce to the Ising model. For the sake of simplicity of the expression of error rates, we continue to treat the divergence model.

Let $\delta = \sqrt{2D(\boldsymbol{\mu}(1), \boldsymbol{\mu}(2); \Sigma)}$ be the Mahalanobis distance, see (6), and let $\mathcal{N}_i$ be a neighborhood consisting $2K$ neighbors of the pixel 0, where K is a fixed positive integer. For example, the first-order neighborhood gives $K = 2$. Our aim is to derive the misclassification rate of the center pixel 0 given features $\boldsymbol{x}_0$, $\boldsymbol{x}_j$ and labels y_j for $j \in \mathcal{N}_0$. Recall that $\widehat{Y}_{\text{Divergence}}$ is an estimate of y_0 by minimizeing (7). Then, the exact error rate $\Pr\{\widehat{Y}_{\text{Divergence}} \neq Y_0\}$ is given by

$$e(\beta_*; \beta, \delta) = \pi_0 \Phi(-\delta/2) + \sum_{k=1}^{K} \pi_k \left\{ \frac{\Phi(-\delta/2 - k\beta_*\delta/K)}{1 + e^{-k\beta\delta^2/K}} + \frac{\Phi(-\delta/2 + k\beta_*\delta/K)}{1 + e^{k\beta\delta^2/K}} \right\} \tag{9}$$

where $\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-t^2/2) dt$ is the cumulative standard Gaussian distribution function, and β_* is the pre-supposed value of the unknown parameter β. Here, π_k gives a probability such that the difference of numbers of pixels labeled 1 and 2 is equal to $k = 0, 1, ..., K$. See Appendix A1.1 for the derivation of the formula (9).

We assume that the prior probability π_0 is less than one. This means that information of neighbors is available for classifying the pixel i. Then, the following properties of the error rate $e(\beta_*; \beta, \delta)$ are shown in A1.2.

P1. $e(0; \beta, \delta) = \Phi(-\delta/2), \quad \lim_{\beta_* \to \infty} e(\beta_*; \beta, \delta) = \pi_0 \Phi(-\delta/2) + \sum_{k=1}^{K} \frac{\pi_k}{1 + e^{k\beta\delta^2/K}}$.

P2. The function $e(\beta_*; \beta, \delta)$ of β_* takes minimum at $\beta_* = \beta$ (Bayes' rule), and the minimum value $e(\beta; \beta, \delta)$ is a monotone decreasing function of δ for a fixed positive constant β.

P3. The function $e(\beta_*; \beta, \delta)$ of β is monotone decreasing for any fixed positive constants β_* and δ.

P4. If the inequality $\beta \geq \frac{1}{\delta^2} \log\left\{ \frac{1 - \Phi(-\delta/2)}{\Phi(-\delta/2)} \right\}$ holds, we have $e(\beta_*; \beta, \delta) < \Phi(-\delta/2)$ for any positive β_*.

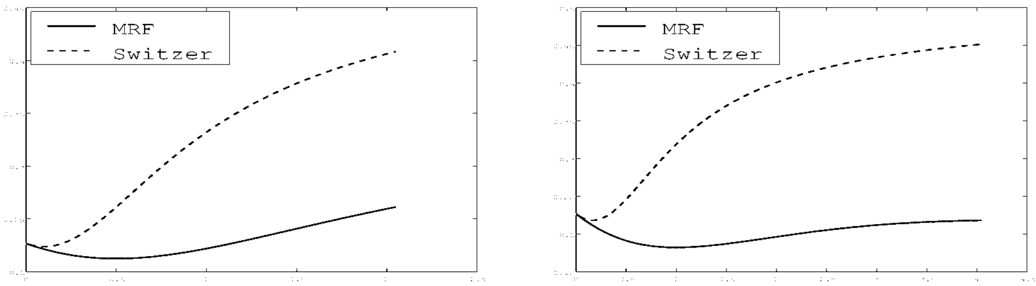
Note that the value $e(0; \beta, \delta) = \Phi(-\delta/2)$ is nothing but the error rate due to Fisher's linear discriminant function in noncontextual case. The asymptotic value $\sum_{k=1}^{K} \pi_k / \left(1 + e^{k\beta\delta^2/K}\right)$ given in P1 is the error rate due to the vote-for-majority rule by neighbors. The property P2 recommends us to use the true parameter β if it is known, and this is quite natural. The property P3 means that the classification becomes efficient when δ and/or β are large. Note that δ is a distance in the feature space, and β is a distance in the geometric space. The property P4 implies that the use of spatial information *always improve noncontextual discrimination* even if the pre-supposed parameter β_* is far from the true value β.

The error rate due to Switzer's method is obtained in the same form of (9) by replacing δ with $\delta_* \equiv \delta \big/ \sqrt{1 + 4\beta_*^2/K}$ appearing in $\Phi(\cdot)$ of the error rate (9).

The comparison of two error rates is illustrated in Figure 1 by two cases of with 4 ($= 2K$) neighbors. The x-axis is corresponding to β_*. It is seen that the divergence model overcomes Switzer's method, and the error rate due to the MRF takes the minimum value at the true values β (P2). The parameters of the figure (a) do not meet the sufficient condition of P4 because $\beta = 0.5 < 0.5455 = \log\left\{\frac{1-\Phi(-\delta/2)}{\Phi(-\delta/2)}\right\}/\delta^2$. Hence, the error rate exceeds $\Phi(-\delta/2)$ for large β_*. Whereas the parameters of the figure (b) meet the condition because the same δ is used. The error rate is always less than $\Phi(-\delta/2)$ for any β_*. Note that the error rate due to the vote-for-majority rule (P1) is given by the asymptotic value of the curve by the MRF.

4. Parameter estimation in MRFs based on divergence

We consider the parameter estimation procedure for the divergence models formulated in Section 3. Suppose that a set of training data $\{(\boldsymbol{x}_i, y_i) \mid i \in \mathcal{D}\}$ is given. Put $\boldsymbol{x} = (\boldsymbol{x}_1^T, ..., \boldsymbol{x}_n^T)^T$ and $\boldsymbol{y} = (y_1, ..., y_n)^T$, and let $\Theta = \{\boldsymbol{\theta}(1), \cdots, \boldsymbol{\theta}(G)\}$ be a set



(a): $\delta = 1.5$, $\beta = 0.5$ (b): $\delta = 1.5, \beta = 1$

Figure 1: Error rates due to MRF and Switzer's method (x-axis: β_*)

of unknown parameters specifying the class-conditional densities (1). We define the logarithms of the conditional likelihood of $\boldsymbol{x}|\boldsymbol{y}$ and the pseudo-likelihood of $\boldsymbol{y}$ as

$$\ell_1(\Theta) = \sum_{i \in \mathcal{D}} \log f(\boldsymbol{x}_i, \boldsymbol{\theta}(y_i)) \quad \text{and} \quad \ell_2(\Theta, \beta) = \sum_{i \in \mathcal{D}} \log p_i(y_i \mid \boldsymbol{y}_{-i}). \tag{10}$$

The parameter Θ appears in both of the likelihoods, and this causes difficulties in parameter estimation. We propose the following three methods for parameter estimation. Among these methods, 4.1 gives the simplest approach, which is widely used for usual setting.

4.1. Estimation of Θ by the conditional distribution

The first method for estimation is based on the conditional distribution $p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$ in (1). The Θ is estimated by maximizing the conditional likelihood as $\widehat{\Theta} = \arg\max_\Theta \ell_1(\Theta)$. The dependency parameter β can be estimated by maximizing the posterior probability of classified test data. Note that the estimation of β through training data is not required.

4.2. Estimation of (Θ, β) by the separate inference

The second method is based on the joint distribution $p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})$ of the training data. Using the estimate $\widehat{\Theta}$ in the subsection 4.1, we maximize the profile log likelihood and define $\widehat{\beta}(\widehat{\Theta}) = \arg\max_\beta \ell_2(\widehat{\Theta}, \beta)$. Then, an estimation equation for β is given as follows:

$$\partial \ell_2(\widehat{\Theta}, \beta)/\partial \beta = \sum_{i \in \mathcal{D}} \sum_{g \in \mathcal{G}} \Delta_i(g)\{\delta_{g,y_i} - p_i(g \mid \boldsymbol{y}_{-i})\} = 0, \tag{11}$$

where $\Delta_i(g)$ is defined by (3), and δ_{gh} is Kronecker's delta. We will take a gradient algorithm for obtaining the optimal value $\widehat{\beta}(\widehat{\Theta})$, and it would be a feasible task.

4.3. Maximization of the sum of the log likelihoods

The last method is to maximize the sum of two log likelihoods. At first, the optimal parameter $\widehat{\Theta}_\beta = \arg\max_\Theta\{\ell_1(\Theta) + \ell_2(\Theta, \beta)\}$ is found for fixed β. Then, the optimal value of β is chosen by $\arg\max_\beta\{\ell_1(\widehat{\Theta}_\beta) + \ell_2(\widehat{\Theta}_\beta, \beta)\}$. An iterative method is developed at Appendix 2 in the Gaussian case as well as a general exponential family case.

5. Applications to real data

Table 1. Error rates (%) of classification results based on homoscedastic /heteroscedastic Gaussian MRFs with two sorts of divergences for the data grss_dfc_0006

Neighborhoods		Gaussian MRFs			
radius	average	Homoscedastic		Heteroscedastic	
r	of $\|\mathcal{N}_i(r)\|$	Ising	Jeffreys	Ising	Jeffreys
0	0.00	8.61	8.61	14.67	14.67
2	3.46	6.44	**6.09**	12.93	**11.44**
3	6.73	6.22	**5.69**	13.09	**10.19**
4	9.72	5.97	**5.35**	12.80	**9.67**
5	15.59	6.35	**5.69**	13.16	**9.39**
6	21.02	6.49	**5.64**	13.16	**9.55**
7	26.27	6.68	**6.13**	12.90	**9.90**
8	33.72	6.70	**6.65**	12.92	**9.76**

Numerals in bold face denote the optimal values in respective rows and two sorts of divergences.

Our method is applied to two multispectral images. In both of the cases, we estimate the parameters by the method described by the subsection 4.1. The neighborhood $\mathcal{N}_i$ is chosen by a set of pixels whose distance from i is less than or equal to a constant r.

5.1. The data set grss_dfc_0006

The benchmark data set grss_dfc_0006, provided by IEEE Geoscience and Remote Sensing society Data Fusion reference database (2001), consists of samples with fifteen feature variables and five land-cover categories on the Feltwell (UK). The training area and the test area consist of 5072 and 5760 pixels respectively. We applied GMRFs with the following four possible cases: homoscedastic or heteroscedastic case; and the Ising model or the divergence model. The optimal classification result is chosen so as to maximize the posterior probability under given radius r of the neighborhoods. We search the optimal combination of the radius $r = 0, 1, ..., 8$ and the spatial-dependency parameter $\beta = 0(.05)10.00$. Table 1 lists error rates of the optimal result given the radius r.

It is seen that both of the spatial models improve the noncontextual classification result ($r = 0$) significantly. In the homoscedastic case, the divergence model gives a similar performance to the Ising model. In the heteroscedastic case, the divergence model is superior to the Ising model in a small degree. See Nishii and Tanaka (1999) for accuracy assessment.

5.2. Simulated multispectral images

Next, an image (a) of Figure 2 with three categories ($G = 3$) is generated. The labels 1, 2 and 3 are corresponding to grey, white and black. The numbers of pixels from the categories are respectively given by 3330, 1371 and 3580. We simulate four-dimensional data ($m = 4$) at each pixel of the image (a) following $N_4(\boldsymbol{\mu}(g), \sigma^2 E_4)$ for $g \in \{1, 2, 3\}$ with $\boldsymbol{\mu}(1) = \mathbf{0}$, $\boldsymbol{\mu}(2) = (1\ 1\ 0\ 0)'/\sqrt{2}$, $\boldsymbol{\mu}(3) = (1.0498\ \ -0.6379\ 0\ 0)'$ and $\sigma^2 = 1, 2$, where E_4 denotes the identity matrix of order 4. The mean vectors are chosen so as to maximize the pseudo-likelihood

(a) (b) (c)

Figure 2: The true image with three categories and the classified results.

(a): The true image of size 91×91. 4-dimensional spectral images are generated over the true image. (b): A classified image by Ising model (Error rate 4.17%). (c): A classified image by the divergence model (Error rate 2.93%).

Table 2. Error rates (%) due to Gaussian MRFs with two variance-covariance matrices $\sigma^2 I$ for $\sigma^2 = 1, 2$.

Neighborhoods		Homoscedastic Gaussian MRFs			
radius	average	$\sigma^2 = 1$		$\sigma^2 = 2$	
r	of $\|\mathcal{N}_i(r)\|$	Ising	Jeffreys	Ising	Jeffreys
0	0.00	41.75	41.75	49.35	49.35
1	3.96	9.56	**9.06**	**19.57**	19.68
$\sqrt{2}$	7.87	3.97	**3.54**	8.92	**7.18**
2	11.78	3.97	**<u>2.33</u>**	7.56	**5.37**
$\sqrt{5}$	19.52	3.20	**2.86**	5.35	**4.88**
3	27.21	3.68	**3.33**	6.41	**6.19**
$\sqrt{10}$	34.86	3.86	**3.83**	6.45	**6.19**

of the true image. Squared Mahalanobis distances are given by $J(1,2) = 1/\sigma^2$, $J(1,3) = 1.5089/\sigma^2$ and $J(2,3) = 1.9264/\sigma^2$.

Table 2 summarizes error rates due to the divergence model with two covariance matrices $\sigma^2 E_4$ for $\sigma^2 = 1, 2$. This table shows that the divergence model shows a better performance than the Ising model. Images (b) and (c) are obtained respectively by Ising model and the divergence model with the radius $r = 2$ and $\beta = 10$ with $\sigma^2 = 1$. It is seen that the divergence model gives a stable classification result.

6. Conclusions

We have considered the MRF specified by the Jeffreys divergence between class-conditional densities, with emphasis on Gaussian distributions. The divergence model is compared with Switzer's smoothing method. The estimation methods of the parameters are also established.

We summarize the features of the divergence model as follows.

- The classification based on the divergence model is a natural extension of Switzer's smoothing method.

- The error rate of our method was obtained in the closed form in a simple setup. It is shown that the use of spatial information always reduces the error rate.
- The divergence model is applied to the two sets of image data. The proposed model shows a better performance than the Ising model.

The divergence model can be defined in an exponential family of probability distributions, which includes most of important distributions. The parameter estimation method to the general family is stated in Appendix 2.

We have proposed three estimation procedures in Section 4. However, the problem which estimation method is efficient for given data still remains. Also model selection is another important problem.

Acknowledgements

The data set grss_dfc_0006 is provided by IEEE GRS-S Data Fusion committee. This research is supported by Grant-in-Aid for Scientific Research (C) 15540123 from Japan Society for the Promotion of Science.

References

Besag, J., 1986, On the statistical analysis of dirty pictures, *J. R. Stat. Soc. B*, **48**, 259-302.

Cressie, N., 1993, *Statistics for Spatial Data*, 2nd edn (New York: John Wiley & Sons).

Chilès, J.P. and Delfiner, P., 1999, *Geostatistics*, (New York: John Wiley & Sons).

Eguchi, S. and Copas, J., 2002, A class of logistic-type discriminant functions. *Biometrika* **89**, 1-22.

Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

IEEE GRSS Data Fusion reference database, 2001, The data set GRSS_DFC_0006. Online. http://www.dfc-grss.org/.

Jain, A.K., Duin, R.P.W., and Mao, J., 2000, Statistical pattern recognition: a review, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **22**, 4-37.

Jeffreys, H., 1946, An invariant form for the prior probability in estimation problems, *Proc. Roy. Soc. London*, ser. A, **186**, 453-461.

Li, S.Z., 2000, Modeling image analysis problems using Markov random fieldsp, *Handbook of Statistics* **20**, (New York: John Wiley & Sons) pp. 1-43.

Mardia, K.V., 1984, Spatial discrimination and classification map, *Commun. Statist.–Theory Meth.*, **13**, 2181–2197. (erratum: *Commun. Statist.–Theory Meth.*, 1987, **16**, p. 3749.)

Marroquin, J., Velasco, F.A., Rivera, M., and Nakamura, M., 2001, Gauss-Markov measure field models for low-level vision. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **23**, 337-348.

McLachlan, G. J., 1992, *Discriminant Analysis and Statistical Pattern Recognition*, (New York: John Wiley & Sons).

Nishii, R., 2003, A Markov random field-based approach to decision level fusion for remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 2316-2319.

Nishii, R. and Tanaka, S., 1999, Accuracy and inaccuracy assessments in land-cover classification, *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 491-498.

Switzer, P., 1980, Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery, *Mathematical Geology*, **12**, 367–376.

Taneja, I.J., 1995, New developments in generalized information measures, Chapter in: *Advances in Imaging and Electron Physics*, edited by P.W. Hawkes, 91, pp. 37-135.

Wilson, R. and Li, C-T., 2003, A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **25**, 42-56.

Appendix 1. Error rates in a local region

A1.1. Derivation of error rates

Let L_1 be a random variable denoting a number of pixels with label 1 in the neighborhood $\mathcal{N}_0$. This implies that the remaining $2K - L_1$ pixels in $\mathcal{N}_0$ are labeled by 2. The label of the center is estimated by (7) for $L_1 = 0, 1, ..., 2K$. By the formula (11), it holds that $\Pr\{Y_1 = 1 \mid L_1 = K + k\} = \Pr\{Y_0 = 2 \mid L_1 = K - k\} = 1/\{1 + \exp(-k\beta\delta^2/K)\}$. Due to the classification rule (7), the following error rates can be obtained by those of the linear discriminant functions

$$\Pr\{\widehat{Y}_{\text{Divergence}} = 2 \mid Y_0 = 1, L_1 = K + k\} = \Phi(-\delta/2 - k\beta_*\delta/K),$$
$$\Pr\{\widehat{Y}_{\text{Divergence}} = 1 \mid Y_0 = 2, L_1 = K + k\} = \Phi(-\delta/2 + k\beta_*\delta/K)$$

for $k = 0, \pm1, ..., \pm K$. Thus, we have the relation: $\Pr\{\widehat{Y}_{\text{Divergence}} \neq Y_0 \mid L_1 = K + k\} = \Pr\{\widehat{Y}_{\text{Divergence}} \neq Y_0 \mid L_1 = K - k\}$. Its value, say $e_k(\beta_*; \beta, \delta)$, is given by

$$e_k(\beta_*; \beta, \delta) = \frac{\Phi(-\delta/2 - k\beta_*\delta/K)}{1 + e^{-k\beta\delta^2/K}} + \frac{\Phi(-\delta/2 + k\beta_*\delta/K)}{1 + e^{k\beta\delta^2/K}} \quad (k = 1, ..., K) \tag{12}$$

Taking the expectation with respect to L_1, we have the error rate presented at (9), where $\pi_k = \Pr\{L_1 = K \pm k\}$ is the prior distribution of L_1 for $k = 0, 1, \ldots, K$.

A1.2. The properties of error rates

The property P1 follows immediately. The property P4 is obtained by the sufficient condition such that $e(0; \beta, \delta) > \lim_{\beta_* \to \infty} e(\beta_*; \beta, \delta)$. Now, we only need to show that the conditional error rates (12) satisfy P2 and P3 since the error rate (9) is given by their convex combination.

Put $k' = k/K$, $\xi = 1/\{1 + \exp(k'\beta\delta^2)\}$, and $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi} = d\,\Phi(x)/dx$. Noting the relation $\phi(-\delta/2 - k'\beta\delta)e^{k'\beta\delta^2} = \phi(-\delta/2 + k'\beta\delta)$, we have the derivatives:

$$\partial e_k(\beta_*; \beta, \delta)/\partial\beta_* = k'\delta\xi\phi(-\delta/2 + k'\beta\delta)\{e^{k'\delta^2(\beta_* - \beta)} - 1\}, \tag{13}$$

$$\begin{aligned}\partial e_k(\beta; \beta, \delta)/\partial\delta = &-\xi\phi(-\delta/2 + k'\beta\delta) \\ &-2k'\beta\delta\xi^2 e^{k'\beta\delta^2}\{\Phi(-\delta/2 + k'\beta\delta) - \Phi(-\delta/2 - k'\beta\delta)\},\end{aligned} \tag{14}$$

$$\partial e_k(\beta_*; \beta, \delta)/\partial\beta = -k'\delta^2\xi^2 e^{k'\beta\delta^2}\{\Phi(-\delta/2 + k'\beta_*\delta) - \Phi(-\delta/2 - k'\beta_*\delta)\}. \tag{15}$$

The derivative (13) implies that the function $e_k(\beta_*;\beta,\delta)$ of β_* takes the minimum value at β. Because of the monotonicity of the function $\Phi(\cdot)$, the derivative (14) implies that the minimum value $e_k(\beta;\beta,\delta)$ is a monotone decreasing function of δ for any positive β. Similarly, the derivative (15) shows the property P3.

Appendix 2. Parameter estimation based on the sum of log likelihoods

Let $\ell(\Theta,\beta)=\ell_1(\Theta)+\ell_2(\Theta,\beta)$ be the sum of two log likelihoods defined by (10). Then, the estimation equations for deriving the optimal parameter $\widehat{\Theta}_\beta=\arg\max_\Theta\{\ell(\Theta,\beta)\}$ for fixed β will be obtained in the following three cases.

A2.1. Homoscedastic GMRFs

Consider Gaussian distributions $N_m(\boldsymbol{\mu}(g),\Sigma)$ with common variance-covariance matrix for g in $\mathcal{G}$. In this case, the sum of two log likelihoods ℓ is given by

$$\ell=-\frac{mn}{2}\log(2\pi)-\frac{n}{2}\log|\Sigma|-\sum_{i\in\mathcal{D}}\Big[D(\boldsymbol{x}_i,\boldsymbol{\mu}(y_i);\Sigma)+\beta\Delta_i(y_i)+\log\Big\{\sum_{h\in\mathcal{G}}\exp\{-\beta\Delta_i(h)\}\Big\}\Big]$$

where $\Delta_i(\cdot)$ is defined by (3) with $J(g,h)=2D(\boldsymbol{\mu}(g),\boldsymbol{\mu}(h);\Sigma)$: the squared Mahalanobis distance. We define the following notations:

$$\begin{array}{ll}\bar r_g(h)=\sum_{i\in\mathcal{D}(g)}r_i(h), & \bar p_+(h)=\sum_{i\in\mathcal{D}}p_i(h\mid\boldsymbol{y}_{-i})\\ \bar r_+(h)=\sum_{i\in\mathcal{D}}r_i(h), & \bar d_+(g,h)=\sum_{i\in\mathcal{D}}r_i(g)p_i(h\mid\boldsymbol{y}_{-i})\end{array}\quad\text{for}\quad g,h\in\mathcal{G},\tag{16}$$

where $\mathcal{D}(g)$ is a set of pixels labeled with g in the training area $\mathcal{D}$, $p_i(\cdot|\cdot)$ and $r_i(\cdot)$ are respectively defined by (2) and (3). The estimating equations for the mean vectors $\boldsymbol{\mu}(g)$ and the covariance matrix Σ can then be obtained by the partial differential equations $\partial\ell/\partial\boldsymbol{\mu}(g)=\mathbf{0}$ and $\partial\ell/\partial\Sigma=O$ as

$$n_g\{\bar{\boldsymbol{x}}(g)-\boldsymbol{\mu}(g)\}-2\beta\sum_{h\in\mathcal{G}}a(g,h)\{\boldsymbol{\mu}(g)-\boldsymbol{\mu}(h)\}=\mathbf{0}\tag{17}$$

$$\Sigma-\frac{1}{n}\sum_{g\in\mathcal{G}}\sum_{i\in\mathcal{D}(g)}\{\boldsymbol{x}_i-\boldsymbol{\mu}(g)\}\{\boldsymbol{x}_i-\boldsymbol{\mu}(g)\}^T-\beta T=O,\tag{18}$$

where $\bar{\boldsymbol{x}}(g)=\sum_{i\in\mathcal{D}(g)}\boldsymbol{x}_i/n_g$, $n_g=|\mathcal{D}(g)|$,

$$T=\frac{2}{n}\sum_{h\in\mathcal{G}}\sum_{h'\in\mathcal{G}}b(h,h')\{\boldsymbol{\mu}(h)-\boldsymbol{\mu}(h')\}\{\boldsymbol{\mu}(h)-\boldsymbol{\mu}(h')\}^T,\tag{19}$$

$$a(g,h)=b(g,h)+b(h,g),\quad b(g,h)=\bar r_g(h)-\bar d_+(h,g).\tag{20}$$

We note that $a(g,h)$ depends on the Mahalanobis distance, so the unknown parameters $\boldsymbol{\mu}(g)$ and Σ are estimated by the iterative procedure. Here, the initial estimates would be recommended to take the sample means and the sample variance-covariance matrix.

A2.2. Heteroscedastic GMRFs

Secondly we investigate the case of Gaussian distributions $N_m(\boldsymbol{\mu}(g),\Sigma(g))$ with different variance-covariance matrices for g in $\mathcal{G}$. Then, the Jeffreys divergence $J(g,h)$ is calculated as

$$\begin{aligned}J(g,h)&=D(\boldsymbol{\mu}(g),\boldsymbol{\mu}(h);\Sigma(g))+D(\boldsymbol{\mu}(g),\boldsymbol{\mu}(h);\Sigma(h))\\&+\frac{1}{2}\mathrm{trace}\Big\{\Sigma(h)^{-1}\Sigma(g)+\Sigma(g)^{-1}\Sigma(h)\Big\}-m.\end{aligned}\tag{21}$$

The estimating equations for $\boldsymbol{\mu}(g)$ and $\Sigma(g)$ with $g \in \mathcal{G}$ are expressed by

$$\begin{aligned} n_g\{\bar{\boldsymbol{x}}(g) - \boldsymbol{\mu}(g)\} &- 2\beta \sum_{h\in\mathcal{G}} b(g,h)\{\boldsymbol{\mu}(g) - \boldsymbol{\mu}(h)\} \\ &- \beta \sum_{h\in\mathcal{G}} b(h,g)\{E_m + \Sigma(g)\Sigma(h)^{-1}\}\{\boldsymbol{\mu}(g) - \boldsymbol{\mu}(h)\} = \mathbf{0}, \end{aligned}$$

$$\begin{aligned} n_g\Sigma(g) &- \sum_{i\in\mathcal{D}(g)} \{\boldsymbol{x}_i - \boldsymbol{\mu}(g)\}\{\boldsymbol{x}_i - \boldsymbol{\mu}(g)\}^T \\ &- \beta \sum_{h\in\mathcal{G}} a(g,h) \Big[\{\boldsymbol{\mu}(g) - \boldsymbol{\mu}(h)\}\{\boldsymbol{\mu}(g) - \boldsymbol{\mu}(h)\}^T + \Sigma(h) - \Sigma(g)\Sigma^{-1}(h)\Sigma(g)\Big] = O \end{aligned}$$

where E_m denotes the identity matrix of order m, and $a(g,h)$ and $b(g,h)$ are defined at (20).

A2.3. Further extension to exponential family

A2.3.1. A natural exponential family and Jeffreys divergence

We proceed to the general case such that the class-conditional densities are of the exponential family. Let the class-conditional densities (1) belong to an exponential family consisting probability densities of the form

$$f(\boldsymbol{x}_*, \boldsymbol{\theta}) = f_0(\boldsymbol{x}_*) \exp\{\boldsymbol{t}(\boldsymbol{x}_*)^T\boldsymbol{\theta} - \kappa(\boldsymbol{\theta})\} \tag{22}$$

for $\boldsymbol{\theta} = \boldsymbol{\theta}(1), \ldots, \boldsymbol{\theta}(G)$, where $\boldsymbol{t}(\boldsymbol{x}_*)$ is a vector of sufficient statistics of the m-dimensional feature vector $\boldsymbol{x}_*$, and $\kappa(\boldsymbol{\theta})$ is the cumulant transform. Then, the Jeffreys divergence $J(g,h)$ in this case is given by

$$J(g,h) = \{\boldsymbol{\eta}(g) - \boldsymbol{\eta}(h)\}^T\{\boldsymbol{\theta}(g) - \boldsymbol{\theta}(h)\} \tag{23}$$

where $\boldsymbol{\eta}(g) = \partial\kappa(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}(g)}$ is a mean vector.

A2.3.2. Parameter estimation methods

We will derive an iterative algorithm for obtaining the pseudo MLE: $\widehat{\Theta}_\beta = \arg\max_\Theta\{l_1(\Theta) + l_2(\Theta,\beta)\}$ for given $\beta \geq 0$, where $l_1(\Theta)$ and $l_2(\Theta,\beta)$ are the log likelihood functions based on the density (22) and the divergence (23) respectively. First, we prepare the following relation owing to partial derivatives for $\boldsymbol{\theta}(g)$ as

$$\partial J(h,h')/\partial\boldsymbol{\theta}(g) = \{\delta_{gh}I(h) - \delta_{gh'}I(h')\}\{\boldsymbol{\theta}(h) - \boldsymbol{\theta}(h')\} + (\delta_{gh} - \delta_{gh'})\{\boldsymbol{\eta}(h) - \boldsymbol{\eta}(h')\}$$

where $I(g) \equiv \partial^2\kappa(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T|_{\boldsymbol{\theta}=\boldsymbol{\theta}(g)} : m \times m$ is the Fisher information and δ_{gh} denotes Kronecker's delta. Next, using the formula in the above and relations $\sum_{g\in\mathcal{G}} r_i(g) = \sum_{g\in\mathcal{G}} p_i(g|\boldsymbol{y}_{-i}) = 1$, we have

$$\begin{aligned} \partial \sum_{i\in\mathcal{D}} \Delta_i(y_i)/\partial\boldsymbol{\theta}(g) &= \sum_{h\in\mathcal{G}} \{\bar{r}_g(h) + \bar{r}_h(g)\}\boldsymbol{\tau}(g,h) \\ \partial \sum_{i\in\mathcal{D}} \log\Big[\sum_{h\in\mathcal{D}} \exp\{-\beta\Delta_i(h)\}\Big]\Big/\partial\boldsymbol{\theta}(g) &= -\beta \sum_{i\in\mathcal{D}}\sum_{h\in\mathcal{G}} \{\bar{d}_+(g,h) + \bar{d}_+(h,g)\}\boldsymbol{\tau}(g,h) \end{aligned}$$

where $\boldsymbol{\tau}(g,h) = I(g)\{\boldsymbol{\theta}(g) - \boldsymbol{\theta}(h)\} + \boldsymbol{\eta}(g) - \boldsymbol{\eta}(h)$, and $\bar{d}_+(g,h)$ is defined by (16) with $J(g,h)$ of (23). Finally, the estimating equation $\partial\{\ell_1(\Theta) + \ell_2(\Theta,\beta)\}/\partial\boldsymbol{\theta}(g) = \mathbf{0}$ is given by

$$n_g\{\bar{\boldsymbol{t}}(g) - \boldsymbol{\eta}(g)\} - \beta \sum_{h\in\mathcal{G}} a(g,h)\boldsymbol{\tau}(g,h) = \mathbf{0}$$

where $a(g,h)$ is defined by (20) with $J(g,h)$ of (23).

List of MHF Preprint Series, Kyushu University

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

MHF2003-1 Mitsuhiro T. NAKAO, Kouji HASHIMOTO & Yoshitaka WATANABE
A numerical method to verify the invertibility of linear elliptic operators with applications to nonlinear problems

MHF2003-2 Masahisa TABATA & Daisuke TAGAMI
Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients

MHF2003-3 Tomohiro ANDO, Sadanori KONISHI & Seiya IMOTO
Adaptive learning machines for nonlinear classification and Bayesian information criteria

MHF2003-4 Kazuhiro YOKOYAMA
On systems of algebraic equations with parametric exponents

MHF2003-5 Masao ISHIKAWA & Masato WAKAYAMA
Applications of Minor Summation Formulas III, Plücker relations, Lattice paths and Pfaffian identities

MHF2003-6 Atsushi SUZUKI & Masahisa TABATA
Finite element matrices in congruent subdomains and their effective use for large-scale computations

MHF2003-7 Setsuo TANIGUCHI
Stochastic oscillatory integrals - asymptotic and exact expressions for quadratic phase functions -

MHF2003-8 Shoki MIYAMOTO & Atsushi YOSHIKAWA
Computable sequences in the Sobolev spaces

MHF2003-9 Toru FUJII & Takashi YANAGAWA
Wavelet based estimate for non-linear and non-stationary auto-regressive model

MHF2003-10 Atsushi YOSHIKAWA
Maple and wave-front tracking — an experiment

MHF2003-11 Masanobu KANEKO
On the local factor of the zeta function of quadratic orders

MHF2003-12 Hidefumi KAWASAKI
Conjugate-set game for a nonlinear programming problem

MHF2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise

MHF2004-2 Rui YAMAGUCHI, Eiko TSUCHIYA & Tomoyuki HIGUCHI
State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors

MHF2004-3 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Cubic pencils and Painlevé Hamiltonians

MHF2004-4 Atsushi KAWAGUCHI, Koji YONEMOTO & Takashi YANAGAWA
Estimating the correlation dimension from a chaotic system with dynamic noise

MHF2004-5 Atsushi KAWAGUCHI, Kentarou KITAMURA, Koji YONEMOTO, Takashi YANAGAWA & Kiyofumi YUMOTO
Detection of auroral breakups using the correlation dimension

MHF2004-6 Ryo IKOTA, Masayasu MIMURA & Tatsuyuki NAKAKI
A methodology for numerical simulations to a singular limit

MHF2004-7 Ryo IKOTA & Eiji YANAGIDA
Stability of stationary interfaces of binary-tree type

MHF2004-8 Yuko ARAKI, Sadanori KONISHI & Seiya IMOTO
Functional discriminant analysis for gene expression data via radial basis expansion

MHF2004-9 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Hypergeometric solutions to the q - Painlevé equations

MHF2004-10 Raimundas VIDŪNAS
Expressions for values of the gamma function

MHF2004-11 Raimundas VIDŪNAS
Transformations of Gauss hypergeometric functions

MHF2004-12 Koji NAKAGAWA & Masakazu SUZUKI
Mathematical knowledge browser

MHF2004-13 Ken-ichi MARUNO, Wen-Xiu MA & Masayuki OIKAWA
Generalized Casorati determinant and Positon-Negaton-Type solutions of the Toda lattice equation

MHF2004-14 Nalini JOSHI, Kenji KAJIWARA & Marta MAZZOCCO
Generating function associated with the determinant formula for the solutions of the Painlevé II equation

MHF2004-15 Kouji HASHIMOTO, Ryohei ABE, Mitsuhiro T. NAKAO & Yoshitaka WATANABE
Numerical verification methods of solutions for nonlinear singularly perturbed problem

MHF2004-16 Ken-ichi MARUNO & Gino BIONDINI
Resonance and web structure in discrete soliton systems: the two-dimensional Toda lattice and its fully discrete and ultra-discrete versions

MHF2004-17 Ryuei NISHII & Shinto EGUCHI
Supervised image classification in Markov random field models with Jeffreys divergence