

Group variable selection via relevance vector machine

Tateishi, Shohei
Graduate School of Mathematics, Kyushu University

Kinjo, Chiaki
Graduate School of Mathematics, Kyushu University

Konishi, Sadanori
Faculty of Mathematics, Kyushu University

<https://hdl.handle.net/2324/19369>

出版情報 : MI Preprint Series. 2011-7, 2011-03-23. 九州大学大学院数理学研究院
バージョン :
権利関係 :



MI Preprint Series

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

Group variable selection via relevance vector machine

**Shohei Tateishi, Chiaki Kinjo &
Sadanori Konishi**

MI 2011-7

(Received March 23, 2011)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Group variable selection via relevance vector machine

Shohei Tateishi*, Chiaki Kinjo* and Sadanori Konishi†

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

† *Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan.*

s-tateishi@math.kyushu-u.ac.jp ma209017@math.kyushu-u.ac.jp konishi@math.chuo-u.ac.jp

Abstract

We consider the problem of variable selection in the case that explanatory variables have some groups. We proposed the extension of relevance vector machine (Tipping, 2001) for variable selection at a group level. In order to estimate a model, we derive a new estimation algorithm along with traditional relevance vector machine. Simulation results demonstrate that our methodology performs well in various situations.

Key Words and Phrases: Group variable selection, Relevance vector machine.

1 Introduction

Variable or feature selection has become one of the most important techniques for selecting a subset of relevant variables when constructing statistical models. The purpose of variable selection are improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying true process generating data (Guyon and Elisseeff, 2003). Traditional methods include stepwise procedures and best subset selection that accepts the best feature or rejects the worst feature on the basis of some model selection criteria, such as Mallows' C_p (Mallows, 1973; 1995), AIC (Akaike, 1973; 1974), BIC (Schwarz, 1978). However, they can cause local optimum results and are very unstable (Brieman, 1996).

As the other variable selection technique, the shrinkage, penalized or regularization method is used. Tibshirani (1996) proposed the lasso, which imposes an L_1 penalty on regression coefficients. The lasso is a special case of bridge estimation (Frank and Friedman, 1993). Owing to the nature of the L_1 penalty, the lasso method encourages sparse solution

which leads both shrinkage and automatic variable selection simultaneously. Therefore, many lasso-type estimation methods have been proposed.

As one of the estimation methods that gives sparse solution like lasso-type, Tipping (2001) proposed a Bayesian estimation procedure called the relevance vector machine (RVM). It is known that RVM yields more sparse solution than support vector machine (SVM; Vapnik, 1998). In addition, RVM is widely used in regression and classification framework because of its manageability.

However, if it is assumed that the statistical model has groups of explanatory variables, ordinary lasso-type estimation methods and RVM can lead to unsatisfactory results because they only select individual explanatory variables rather than explanatory factors. In order to overcome this problem, we propose an extension of RVM for selecting these groups effectively and derive new update algorithm. Our proposed method does not require choosing regularization parameters that adjusts the degree of the regularization, whereas this is absolutely imperative for lasso-type regularization methods. The proposed modeling procedure is investigated by analyzing Monte Carlo simulations including regression and classification frameworks. The results demonstrate the effectiveness of the proposed method in terms of prediction accuracy.

This paper is organized as follows. Section 2 describes ordinary RVM setting for linear regression model. Therefore, we present new RVM estimation procedure for regression model in Section 3, and extend the framework to classification case in Section 4. In Section 5 we investigate the performance of our modeling techniques through Monte Carlo simulations. Some concluding remarks are presented in Section 6.

2 RVM regression

Suppose that we have n independent observations $\{(y_i, \mathbf{x}_i); i = 1, 2, \dots, n\}$, where y_i is a random response variable and $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})^T$ is a p -dimensional explanatory variable vector. If \mathbf{x} has G groups of variables, we can rewrite \mathbf{x}_i as $\mathbf{x}_i = (\mathbf{x}_{i,1}^T, \dots, \mathbf{x}_{i,G}^T)^T$, where $\mathbf{x}_{i,g}$ is a p_g -dimensional explanatory variable vector corresponding to the g th group ($g = 1, \dots, G$), that is, $\mathbf{x}_{i,g} = (x_{i,g,1}, \dots, x_{i,g,p_g})^T$ and $\sum_{g=1}^G p_g = p$. We consider the

regression model

$$y_i = \beta_0 + \sum_{g=1}^G \mathbf{x}_{i,g}^T \boldsymbol{\beta}_g + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where β_0 is an unknown intercept and $\boldsymbol{\beta}_g$ is a p_g -dimensional unknown coefficient vector. If ϵ_i are independently, normally distributed with mean zero and variance ρ^{-1} , the linear regression model (1) has probability density function

$$f(y_i|\boldsymbol{\beta}, \rho) = \frac{1}{\sqrt{2\pi\rho^{-1}}} \exp \left[-\frac{\{y_i - (\beta_0 + \sum_{g=1}^G \mathbf{x}_{i,g}^T \boldsymbol{\beta}_g)\}^2}{2\rho^{-1}} \right], \quad i = 1, \dots, n. \quad (2)$$

Next we suppose that the $(p+1)$ -dimensional coefficient vector $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_G^T)^T$ has Gaussian prior density

$$\pi(\boldsymbol{\beta}|\boldsymbol{\alpha}) = (2\pi)^{-\frac{p+1}{2}} |A|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \boldsymbol{\beta}^T A \boldsymbol{\beta} \right), \quad (3)$$

where $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_p)^T$ is a $(p+1)$ hyperparameter vector and $A = \text{diag}(\alpha_0, \dots, \alpha_p)$. The posterior distribution for $\boldsymbol{\beta}$ given the data $\mathbf{y} = (y_1, \dots, y_n)^T$ is defined by

$$\pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \rho) = \frac{f(\mathbf{y}|\boldsymbol{\beta}, \rho) \pi(\boldsymbol{\beta}|\boldsymbol{\alpha})}{\int f(\mathbf{y}|\boldsymbol{\beta}, \rho) \pi(\boldsymbol{\beta}|\boldsymbol{\alpha}) d\boldsymbol{\beta}}, \quad (4)$$

where $f(\mathbf{y}|\boldsymbol{\beta}, \alpha) = \prod_{i=1}^n f(y_i|\boldsymbol{\beta}, \alpha)$ and then, we see that the posterior distribution for $\boldsymbol{\beta}$ has Gaussian density

$$\pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \rho) = (2\pi)^{-\frac{p+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \right\},$$

where the posterior covariance matrix and mean vector are respectively

$$\Sigma = (\rho X^T X + A)^{-1}, \quad \boldsymbol{\mu} = \rho \Sigma X^T \mathbf{y}, \quad (5)$$

where $X = (\mathbf{1}_n, X_1, \dots, X_G)$, $X_g = (\mathbf{x}_{1,g}, \dots, \mathbf{x}_{n,g})^T$ ($g = 1, \dots, G$) and $\mathbf{1}_n$ denotes an n -vector whose elements are all ones.

The values of hyperparameters $\boldsymbol{\alpha}$ and ρ are determined by using expectation-maximization (EM) updates, treating the coefficients as the hidden variables and maximizing the expected complete log-likelihood function

$$E_{\pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \rho)} [\log f(\mathbf{y}|\boldsymbol{\beta}, \rho) \pi(\boldsymbol{\beta}|\boldsymbol{\alpha})], \quad (6)$$

where $E_{\pi(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\alpha},\rho)}[\cdot]$ denotes an expectation with respect to the posterior distribution $\pi(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\alpha},\rho)$ over the coefficients given the data and hidden variables. Setting the derivatives of (6) to zero, we obtain estimators of $\boldsymbol{\alpha}, \rho$ given by

$$\hat{\alpha}_j = \frac{1}{\Sigma_{jj} + \mu_j^2}, \quad (\hat{\rho}^{-1})^{\text{new}} = \frac{\|\mathbf{y} - X\boldsymbol{\mu}\|^2 + (\rho^{-1})^{\text{old}} \sum_k \eta_k}{n}, \quad j = 0, \dots, p. \quad (7)$$

where $\eta_j = 1 - \alpha_j \Sigma_{jj}$, μ_j is the $(j+1)$ th element of $\boldsymbol{\mu}$ and Σ_{jj} is the $(j+1)$ th diagonal element of Σ and $\|\cdot\|$ is the Euclidian norm. Because these estimators depend on each other, re-estimation of (5) and (7) is needed. The technique for estimation by sequential computation based on the maximizing marginal likelihood is known as relevance vector machine (RVM; Tipping, 2001) and encourages high sparsity. As the optimization of the hyperparameters progresses, many α s tend towards infinity, so that most coefficients to be estimated are approaching zero. However, when ordinary RVM is directly applied to a regression model (1), individual explanatory variables can be selected instead of groups of predictors.

In order to do variable selection adequately, we propose replacing the traditional Gaussian prior (3) with

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}) = (2\pi)^{-\frac{n}{2}} |\Gamma|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T \Gamma \boldsymbol{\beta}\right), \quad (8)$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_G)^T$ is a $(p+1)$ -dimensional hyperparameter vector, $\boldsymbol{\gamma}_g = (\gamma_g, \dots, \gamma_g)^T$ ($g = 1, \dots, G$) is a p_g -dimensional hyperparameter vector, and $\Gamma = \text{diag}(\boldsymbol{\gamma})$. It becomes possible to select each variable group by grouped hyperparameters in prior (8). In other words, our proposed method encourages sparsity at the group level instead of at the element level.

For likelihood function (2) and prior density (8), we see that the posterior distribution for $\boldsymbol{\beta}$ has Gaussian density

$$\pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\gamma}, \rho) = (2\pi)^{-\frac{p+1}{2}} |\Lambda|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\xi})^T \Lambda^{-1}(\boldsymbol{\beta} - \boldsymbol{\xi})\right\},$$

where the posterior covariance matrix and mean vector are respectively

$$\Lambda = (\rho X^T X + \Gamma)^{-1}, \quad \boldsymbol{\xi} = \rho \Lambda X^T \mathbf{y}. \quad (9)$$

Thus, maximizing the expected complete log-likelihood function leads to the following

update:

$$\hat{\gamma}_0 = \frac{1}{\Lambda_{11} + \xi_1^2}, \quad \hat{\gamma}_g = \frac{p_g}{\sum_{j=1+S_{g-1}}^{S_g} (\Lambda_{jj} + \xi_j^2)}, \quad (g = 1, \dots, G), \quad (10)$$

$$(\hat{\rho}^{-1})^{\text{new}} = \frac{\|\mathbf{y} - X\boldsymbol{\xi}\|^2 + (\rho^{-1})^{\text{old}} \sum_k \zeta_k}{n}, \quad (11)$$

where $S_0 = 1$, $S_k = 1 + \sum_{j=1}^k p_j$ ($k = 1, \dots, G$), $\zeta_j = 1 - \gamma_j \Lambda_{jj}$, ξ_j is $(j+1)$ th element of $\boldsymbol{\xi}$ and Λ_{jj} is the $(j+1)$ th diagonal element of Λ .

3 RVM classification

Suppose that we have n independent observations $\{(y_i, \mathbf{x}_i); i = 1, 2, \dots, n\}$, where y_i is a binary response variable (i.e., $y_i \in \{0, 1\}$) and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is a p -dimensional explanatory variable vector consisting of G groups, as described in the previous section on regression. We consider the problem of constructing logistic models. In the logistic model, we assume that

$$\Pr(Y_i = 1|\mathbf{x}_i) = p(\mathbf{x}_i), \quad \Pr(Y_i = 0|\mathbf{x}_i) = 1 - p(\mathbf{x}_i), \quad (12)$$

where Y_i is regarded as a random variable distributed according to the Bernoulli distribution in the form

$$f(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = p(\mathbf{x}_i)^{y_i} \{1 - p(\mathbf{x}_i)\}^{1-y_i}. \quad (13)$$

The logistic model further assumes that

$$\log \left\{ \frac{p(\mathbf{x}_i)}{1 - p(\mathbf{x}_i)} \right\} = \beta_0 + \sum_{g=1}^G \mathbf{x}_{i,g}^T \boldsymbol{\beta}_g + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (14)$$

Unlike regression framework, closed form expressions for both the posterior $\pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha})$ and marginal likelihood $p(\mathbf{y}|\boldsymbol{\alpha})$ are precluded. We therefore employ Taylor expansion over centered at $\boldsymbol{\beta}_M$:

$$\begin{aligned} \log \pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}) &\approx \log \pi(\boldsymbol{\beta}_M|\mathbf{y}, \boldsymbol{\alpha}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_M)^T \frac{\partial}{\partial \boldsymbol{\beta}} \log \pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}) \Big|_{\boldsymbol{\beta}_M} \\ &\quad - \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_M)^T \left\{ -\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log \pi(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}) \Big|_{\boldsymbol{\beta}_M} \right\} (\boldsymbol{\beta} - \boldsymbol{\beta}_M). \end{aligned} \quad (15)$$

Here, when β_M is assumed to be the posterior mode, that is, the solution of $\partial \log \pi(\beta|\mathbf{y}, \alpha) / \partial \beta = \mathbf{0}$, the second term of (15) becomes equal to zero. Then, we approximate the posterior distribution by Gaussian distribution.

Using the fact that $\partial \log \pi(\beta|\mathbf{y}, \alpha) / \partial \beta = \mathbf{0}$,

$$\begin{aligned} \left. \frac{\partial}{\partial \beta} \log \pi(\beta|\mathbf{y}, \alpha) \right|_{\beta_M} &= \sum_{i=1}^n \left[y_i - \frac{\exp(\beta_M^T \phi(\mathbf{x}_i))}{1 + \exp(\beta_M^T \phi(\mathbf{x}_i))} \right] \phi(\mathbf{x}_i) - A\beta_M \\ &= X^T(\mathbf{y} - \mathbf{p}) - A\beta_M = \mathbf{0} \\ \Leftrightarrow \beta_M &= A^{-1}X^T(\mathbf{y} - \mathbf{p}), \end{aligned} \quad (16)$$

where $\mathbf{p} = (p(\mathbf{x}_1), \dots, p(\mathbf{x}_n))^T$.

The Hessian of (15) is

$$\begin{aligned} \left. \frac{\partial^2}{\partial \beta \partial \beta^T} \log \pi(\beta|\mathbf{y}, \alpha) \right|_{\beta_M} &= - \sum_{i=1}^n \left[\frac{\exp(\beta_M^T \mathbf{x}_i)}{\{1 + \exp(\beta_M^T \mathbf{x}_i)\}^2} \right] \mathbf{x}_i \mathbf{x}_i^T - A, \\ &= -X^T P X - A, \end{aligned} \quad (17)$$

where P denote an $n \times n$ diagonal matrix with $P_{ii} = p(\mathbf{x}_i)\{1 - p(\mathbf{x}_i)\}$ for the i th diagonal element. This is then inverted to give the covariance matrix Σ for a Gaussian approximation to the posterior, that is,

$$\Sigma = (X^T P X + A)^{-1} \quad (18)$$

As a result, the posterior distribution of coefficients β is obtained using the proposed Gaussian prior (8) as follows:

$$\pi(\beta|\mathbf{y}, \gamma) \approx (2\pi)^{-\frac{n}{2}} |\Lambda|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\beta - \xi)^T \Lambda^{-1} (\beta - \xi) \right\}, \quad (19)$$

where the posterior covariance matrix and mean vector are respectively

$$\Lambda = (X^T P X + \Gamma)^{-1}, \quad \xi = A^{-1}X^T(\mathbf{y} - \mathbf{p}). \quad (20)$$

The hyperparameter vector γ is updated using (10) in an analogous fashion to the regression case.

4 Numerical examples

In this section, we describe Monte Carlo simulations conducted to investigate the effectiveness of our proposed regression method and the classification modeling procedures. We use a simulation setting that is similar to that of Yuan and Lin (2006).

4.1 Simulation study for regression

We compared the performance of the proposed method (Group RVM) with those of group lasso (Yuan and Lin, 2006), group bridge (Huang *et al.*, 2009; Breheny and Huang, 2009), ordinary RVM and ordinary least-squares estimation of the full model. $n = 100$ observations were collected from the true regression model $Y = u + \epsilon$ for each simulation data set. We considered the following four cases for the true regression models.

- (a) 15 random variables X_1, \dots, X_{15} were first simulated according to a centered multivariate normal distribution with covariance $\tau^{|i-j|}$ between X_i and X_j . Then, X_i is trichotomized as 0, 1 or 2 according to whether it is smaller than $\Phi^{-1}(\frac{1}{3})$, larger than $\Phi^{-1}(\frac{2}{3})$, or in between. The response variable Y was simulated from the true model

$$\begin{aligned} u = & 2I(X_1 = 1) - 1.5I(X_1 = 0) + 2I(X_3 = 1) + 1.5I(X_3 = 0) \\ & - 2I(X_5 = 1) + 1.5I(X_5 = 0), \end{aligned}$$

where $I(\cdot)$ is the indicator function, and the noise ϵ is normally distributed with mean 0 and variance 1.5^2 .

- (b) Both main effects and second-order interactions were considered. Four categorical factors X_1, X_2, X_3 and X_4 were first generated in the same manner as in model (a). The true regression model is

$$\begin{aligned} u = & 5I(X_1 = 1) + 4I(X_1 = 0) - 5I(X_2 = 1) - 4I(X_2 = 0) + 2I(X_1 = 1, X_2 = 1) \\ & - 3I(X_1 = 1, X_2 = 0) - 2I(X_1 = 0, X_2 = 1) + 3I(X_1 = 0, X_2 = 0), \end{aligned}$$

with mean 0 and variance 2^2 .

- (c) 17 random variables Z_1, \dots, Z_{16} and W were independently generated from a standard normal distribution. The covariates are then defined as $X_i = (Z_i + W)/\sqrt{2}$. The response variable follows

$$u = 2X_3^3 + 2X_3^2 + 2X_3 + 2X_6^3 - 2X_6^2 - 4X_6,$$

where $\epsilon \sim N(0, 2^2)$.

(d) Covariates X_1, \dots, X_{10} were generated in the same manner as in model (c). Then, these 10 covariates X_{11}, \dots, X_{20} were trichotomized as in the first two models. The true regression model is given by

$$u = 2X_3^3 + 2X_3^2 + 2X_3 + \frac{2}{3}X_6^3 - 2X_6^2 + \frac{4}{3}X_6 + 2I(X_{11} = 1) - 2I(X_{11} = 0),$$

where $\epsilon \sim N(0, 2^2)$.

We performed 200 repetitions, then calculated averages of mean squared errors (MSE) defined by $\text{MSE} = \sum_{\alpha}^n \{u_{\alpha} - \hat{y}_{\alpha}\}^2 / n$ and the standard deviations to assess the goodness of fit. In order to choose the optimal smoothing parameter, we use a C_p -type criterion (Yuan and Lin, 2006) for the group lasso and AIC, GCV, BIC (Huang *et al.*, 2009) for group bridge. Table 1 and 2 displays simulation results with mean of number of factors (or interactions) selected, their standard deviations, MSE, and standard deviations of MSE for (a) to (d). In all cases, our proposed modeling procedure minimized the MSE, thus improving the accuracy of prediction. The proposed method tend to select the variables included in a true model whereas group lasso tend to include extra variables.

Table 1: Comparison of results for regression simulations ($\tau = 0$).

| Case | Method | Criterion | Number of groups | MSE |
|------|--------------|-----------|------------------|-------------|
| (a) | Group RVM | – | 2.90 (0.33) | 0.36 (0.34) |
| | Group lasso | C_p | 9.67 (2.90) | 0.43 (0.19) |
| | Group bridge | AIC | 10.45 (1.90) | 0.58 (0.21) |
| | Group bridge | GCV | 7.25 (1.79) | 0.42 (0.18) |
| | Group bridge | BIC | 7.24 (1.75) | 0.43 (0.19) |
| | RVM | – | 8.46 (1.80) | 0.45 (0.20) |
| | Least square | – | 15 (0) | 0.82 (0.24) |
| | True model | – | 3 | 0 |
| (b) | Group RVM | – | 4.09 (0.97) | 0.45 (0.22) |
| | Group lasso | C_p | 7.48 (1.75) | 0.55 (0.24) |
| | Group bridge | AIC | 8.68 (1.05) | 1.07 (0.34) |
| | Group bridge | GCV | 5.35 (1.19) | 0.65 (0.26) |
| | Group bridge | BIC | 6.82 (1.37) | 0.82 (0.30) |
| | RVM | – | 6.82 (1.20) | 0.55 (0.22) |
| | Least square | – | 10 (0) | 1.06 (0.28) |
| | True model | – | 3 | 0 |
| (c) | Group RVM | – | 2.51 (0.69) | 0.34 (0.22) |
| | Group lasso | C_p | 13.26 (2.78) | 1.06 (0.40) |
| | Group bridge | AIC | 13.77 (1.55) | 1.48 (0.38) |
| | Group bridge | GCV | 7.61 (2.02) | 0.78 (0.29) |
| | Group bridge | BIC | 9.91 (2.23) | 1.02 (0.35) |
| | RVM | – | 3.17 (1.11) | 0.55 (0.30) |
| | Least square | – | 16 (0) | 1.93 (0.37) |
| | True model | – | 2 | 0 |
| (d) | Group RVM | – | 5.73 (1.57) | 0.57 (0.27) |
| | Group lasso | C_p | 16.18 (3.51) | 1.17 (0.46) |
| | Group bridge | AIC | 16.23 (2.09) | 1.53 (0.41) |
| | Group bridge | GCV | 9.61 (2.53) | 0.93 (0.36) |
| | Group bridge | BIC | 11.68 (2.48) | 1.11 (0.38) |
| | RVM | – | 9.35 (2.71) | 0.83 (0.33) |
| | Least square | – | 20 (0) | 2.00 (0.42) |
| | True model | – | 3 | 0 |

Table 2: Comparison of results for regression simulations ($\tau = 0.5$).

| Case | Method | Criterion | Number of groups | MSE |
|------|--------------|-----------|------------------|-------------|
| (a) | Group RVM | – | 2.97 (0.29) | 0.31 (0.26) |
| | Group lasso | C_p | 9.53 (2.89) | 0.40 (0.18) |
| | Group bridge | AIC | 10.09 (1.91) | 0.54 (0.21) |
| | Group bridge | GCV | 6.92 (1.89) | 0.39 (0.17) |
| | Group bridge | BIC | 6.97 (1.86) | 0.39 (0.18) |
| | RVM | – | 8.07 (1.84) | 0.41 (0.18) |
| | Least square | – | 15 (0) | 0.78 (0.23) |
| | True model | – | 3 | 0 |
| (b) | Group RVM | – | 3.92 (0.38) | 0.38 (0.19) |
| | Group lasso | C_p | 7.67 (1.94) | 0.60 (0.26) |
| | Group bridge | AIC | 8.53 (1.19) | 0.88 (0.31) |
| | Group bridge | GCV | 5.28 (1.33) | 0.55 (0.23) |
| | Group bridge | BIC | 6.16 (1.39) | 0.63 (0.25) |
| | RVM | – | 6.58 (1.16) | 0.57 (0.23) |
| | Least square | – | 10 (0) | 1.16 (0.29) |
| | True model | – | 3 | 0 |
| (c) | Group RVM | – | 2.53 (0.74) | 0.33 (0.20) |
| | Group lasso | C_p | 13.32 (2.61) | 1.05 (0.41) |
| | Group bridge | AIC | 13.68 (1.63) | 1.46 (0.40) |
| | Group bridge | GCV | 7.24 (2.25) | 0.79 (0.31) |
| | Group bridge | BIC | 9.59 (2.24) | 1.01 (0.35) |
| | RVM | – | 3.43 (1.27) | 0.55 (0.27) |
| | Least square | – | 16 (0) | 1.90 (0.38) |
| | True model | – | 2 | 0 |
| (d) | Group RVM | – | 5.79 (1.64) | 0.58 (0.25) |
| | Group lasso | C_p | 16.39 (3.35) | 1.21 (0.45) |
| | Group bridge | AIC | 16.49 (2.11) | 1.55 (0.40) |
| | Group bridge | GCV | 9.09 (2.37) | 0.91 (0.33) |
| | Group bridge | BIC | 11.37 (2.52) | 1.11 (0.37) |
| | RVM | – | 9.20 (2.46) | 0.82 (0.32) |
| | Least square | – | 20 (0) | 2.01 (0.40) |
| | True model | – | 3 | 0 |

4.2 Simulation study for classification

We compared the performance of the proposed method with those of group lasso for the logistic regression (Meier *et al.*, 2008), group bridge, ordinary RVM and ordinary least squares estimation of the full model. The simulation data were collected from the true model $Y = 1/\{1 - \exp(u)\}$, and then Y was dichotomized as 0 or 1 if it was smaller than 0.5 or not, respectively. $n = 400$ observations were used in each simulation run. We considered the following four cases for the true models as follows.

- (a) 15 random variables X_1, \dots, X_{15} were first simulated according to a centered multivariate normal distribution with covariance $\tau^{|i-j|}$ between X_i and X_j . Then X_i is divided into four categories as 0, 1, 2, and 3 using the quartiles of the standard normal distribution. The response variable Y was simulated from true model

$$\begin{aligned} u = & 2I(X_1 = 0) + 4I(X_1 = 1) + 2I(X_1 = 2) \\ & - 2I(X_3 = 0) + 4I(X_3 = 1) - 2I(X_3 = 2) \\ & + 2I(X_5 = 0) - 4I(X_5 = 1) + 2I(X_5 = 2), \end{aligned}$$

where $I(\cdot)$ is the indicator function.

- (b) Both main effects and second-order interactions were considered. Four categorical factors X_1, X_2, X_3 and X_4 were first generated as in model (a). The true regression model is

$$\begin{aligned} u = & 2I(X_1 = 0) - 4I(X_1 = 1) - 2I(X_1 = 2) \\ & + 4I(X_2 = 0) - 2I(X_2 = 1) - I(X_2 = 2) \\ & - 1.5I(X_1 = 0, X_2 = 0) + 2I(X_1 = 0, X_2 = 1) + 2.5I(X_1 = 0, X_2 = 2) \\ & + 1.5I(X_1 = 1, X_2 = 0) + 1.8I(X_1 = 1, X_2 = 1) + 2I(X_1 = 1, X_2 = 2) \\ & + 2.2I(X_1 = 2, X_2 = 0) + 2.4I(X_1 = 2, X_2 = 1) - 2.6I(X_1 = 2, X_2 = 2). \end{aligned}$$

- (c) 21 random variables Z_1, \dots, Z_{20} and W were independently generated from a standard normal distribution. The covariates were then defined as $X_i = (Z_i + W)/\sqrt{2}$. The response variable follows

$$u = 2X_2^4 + 3X_2^3 + 4X_2^2 + 2X_2 - 3X_3^4 - 4X_3^3 - 2X_3^2 - 3X_3.$$

(d) Covariates X_1, \dots, X_{10} were generated in the same fashion as in model (c). Then, the 10 covariates X_{11}, \dots, X_{20} were trichotomized as in models (a) and (b). The true regression model is given by

$$\begin{aligned} u = & 2X_2^4 + 3X_2^3 + 4X_2^2 + 2X_2 - 3X_3^4 - 4X_3^3 - 2X_3^2 - 3X_3. \\ & + 2I(X_{11} = 0) - 2I(X_{11} = 1) + 2I(X_{11} = 2). \end{aligned}$$

We performed 200 repetitions, then calculated averages of test error rates (ERR) and the standard deviations for $n/2$ test data to assess the goodness of fit. The smoothing parameter of group lasso is selected by using two-fold cross validation (CV) based on log-likelihood (Meier *et al.*, 2008) and that of group bridge is chosen by AIC, GCV and BIC (Breheny and Huang, 2009). Table 3 and 4 display simulation results of the mean of number of factors (or interactions) selected, their standard deviations, ERR, and the standard deviations of ERR for (a) to (d). In all cases, our proposed modeling procedure minimized the ERR, thus improving the accuracy of prediction. The proposed method tend to select the variables included in a true model whereas group lasso tend to include extra variables.

Table 3: Comparison of results for classification simulations ($\tau = 0$).

| Case | Method | Criterion | Number of groups | ERR (%) |
|------|--------------|-----------|------------------|--------------|
| (a) | Group RVM | – | 3.01 (0.07) | 13.22 (2.42) |
| | Group lasso | CV | 13.20 (1.40) | 13.57 (2.48) |
| | Group bridge | AIC | 7.61 (1.85) | 13.66 (2.51) |
| | Group bridge | GCV | 12.05 (2.64) | 14.22 (2.64) |
| | Group bridge | BIC | 3.61 (0.76) | 13.41 (2.49) |
| | RVM | – | 6.88 (1.69) | 13.56 (2.39) |
| | Least square | – | 15 (0) | 15.08 (2.66) |
| | True model | – | 3 | 0 |
| (b) | Group RVM | – | 3.81 (0.79) | 10.99 (2.35) |
| | Group lasso | CV | 8.47 (1.03) | 11.23 (2.44) |
| | Group bridge | AIC | 5.29 (1.42) | 11.48 (2.54) |
| | Group bridge | GCV | 8.33 (1.08) | 13.18 (2.96) |
| | Group bridge | BIC | 3.39 (0.60) | 10.95 (2.35) |
| | RVM | – | 7.66 (0.94) | 11.85 (2.65) |
| | Least square | – | 10 (0) | 12.77 (2.71) |
| | True model | – | 3 | 0 |
| (c) | Group RVM | – | 2.23 (0.50) | 10.15 (4.72) |
| | Group lasso | CV | 14.81 (2.91) | 10.99 (4.68) |
| | Group bridge | AIC | 6.18 (2.78) | 11.80 (4.77) |
| | Group bridge | GCV | 7.27 (3.43) | 12.03 (4.89) |
| | Group bridge | BIC | 2.68 (0.84) | 11.26 (4.47) |
| | RVM | – | 3.05 (1.18) | 10.77 (4.82) |
| | Least square | – | 16 (0) | 14.14 (4.96) |
| | True model | – | 2 | 0 |
| (d) | Group RVM | – | 3.21 (0.58) | 9.29 (3.68) |
| | Group lasso | CV | 16.23 (2.58) | 10.72 (3.91) |
| | Group bridge | AIC | 7.79 (2.83) | 10.19 (3.96) |
| | Group bridge | GCV | 9.82 (3.92) | 10.71 (4.20) |
| | Group bridge | BIC | 3.68 (0.89) | 9.40 (3.69) |
| | RVM | – | 6.42 (2.11) | 10.23 (3.88) |
| | Least square | – | 20 (0) | 13.25 (4.30) |
| | True model | – | 3 | 0 |

Table 4: Comparison of results for classification simulations ($\tau = 0.5$).

| Case | Method | Criterion | Number of groups | ERR (%) |
|------|--------------|-----------|------------------|--------------|
| (a) | Group RVM | – | 3.03 (0.16) | 13.10 (2.55) |
| | Group lasso | CV | 13.04 (1.57) | 13.32 (2.59) |
| | Group bridge | AIC | 7.49 (1.86) | 13.26 (2.51) |
| | Group bridge | GCV | 11.84 (2.13) | 13.93 (2.79) |
| | Group bridge | BIC | 3.55 (0.73) | 13.21 (2.51) |
| | RVM | – | 6.60 (1.64) | 13.15 (2.41) |
| | Least square | – | 15 (0) | 14.90 (2.75) |
| | True model | – | 3 | 0 |
| (b) | Group RVM | – | 4.04 (0.95) | 13.48 (2.33) |
| | Group lasso | CV | 8.48 (1.19) | 13.66 (2.38) |
| | Group bridge | AIC | 5.55 (1.58) | 14.20 (2.61) |
| | Group bridge | GCV | 8.23 (1.24) | 15.26 (2.56) |
| | Group bridge | BIC | 2.31 (0.50) | 13.56 (2.44) |
| | RVM | – | 7.46 (0.96) | 14.28 (2.45) |
| | Least square | – | 10 (0) | 15.15 (2.60) |
| | True model | – | 3 | 0 |
| (c) | Group RVM | – | 2.15 (0.47) | 10.93 (4.88) |
| | Group lasso | CV | 15.17 (2.95) | 12.42 (4.89) |
| | Group bridge | AIC | 6.84 (2.92) | 13.23 (8.97) |
| | Group bridge | GCV | 9.42 (4.03) | 13.82 (9.09) |
| | Group bridge | BIC | 2.79 (1.34) | 12.39 (9.04) |
| | RVM | – | 3.17(1.40) | 11.82 (4.88) |
| | Least square | – | 16 (0) | 15.48 (5.42) |
| | True model | – | 2 | 0 |
| (d) | Group RVM | – | 3.22 (0.61) | 10.24 (3.89) |
| | Group lasso | CV | 16.72 (2.40) | 11.76 (3.95) |
| | Group bridge | AIC | 7.85 (2.58) | 11.16 (3.96) |
| | Group bridge | GCV | 10.62 (3.63) | 11.75 (4.21) |
| | Group bridge | BIC | 3.60 (0.91) | 10.34 (3.88) |
| | RVM | – | 6.16(2.00) | 11.15 (3.97) |
| | Least square | – | 20 (0) | 14.25 (4.46) |
| | True model | – | 3 | 0 |

5 Concluding remarks

We have proposed a group variable selection procedure along with the technique of RVM. When we apply our proposed method to the statistical model which has the groups of explanatory variables, proper estimation and variable selection at a group level are conducted. Our proposed method does not require choosing regularization parameter, whereas it is necessary for lasso-type regularization methods. The effectiveness of the proposed modeling procedures has been shown through various numerical examples.

References

- [1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Proc. 2nd Int. Symp. Information Theory (Petrov, B. N. and Csaki, F. eds.), Akademiai Kiado, Budapest, 267–281.
- [2] Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Contr. AC.19, 716–723.
- [3] Breheny, P. and Huang, J. (2009). Penalized methods for bi-level variable selection. *Statistics and its interface* **2**, 369–380.
- [4] Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Ann. Stat.*, **24**, 2350–2383.
- [5] Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–148.
- [6] Guyon, I. and Elisseeff, A. (2003) An introduction to variable and feature selection, *Journal of Machine Learning Research*, **3**, 1157–1182.
- [7] Huang, J., Ma, S., Xie, H. and Zhang, C. (2009). A group bridge approach for variable selection. *Biometrika* **96**, 339–355.
- [8] Mallows, C. L. (1973). Some comments on c_p . *Technometrics* **15**, 661–675.
- [9] Mallows, C. L. (1995). More comments on c_p . *Technometrics* **37**, 362–372.

- [10] Meier, L., van de Geer, S. and Bühlmann, P. (2008). The Group Lasso for logistic regression. *Journal of the Royal Statistical Society Series B* **70**, 53–71.
- [11] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.
- [12] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B* **58**, 267–288.
- [13] Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* **1**, 211–244.
- [14] Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York.
- [15] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B* **68**, 49–67.

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 initial 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 infinite 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 HIRANOUCI & Yuichiro TAGUCHI
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 q -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 MIIKE & 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
- MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation
- MI2010-8 Takahiro ITO
Abstract collision systems on groups
- MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

- MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models
- MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems
- MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow
- MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates
- MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight
- MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency
- MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres
- MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space
- MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar bose field
- MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling
- MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

- MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test
- MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs
- MI2010-23 Takahiro ITO, Mitsuhiro FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups
- MI2010-24 Toshimitsu TAKAESU
A Hardy's Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra
- MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators
- MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling
- MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time
- MI2010-28 Jun-ichi INOBUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves
- MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves
- MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the q -VBS state of an integer spin model
- MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions

- MI2010-32 Nobutaka NAKAZONO
Hypergeometric τ functions of the q -Painlevé systems of type $(A_2 + A_1)^{(1)}$
- MI2010-33 Yoshiyuki KAGEI
Global existence of solutions to the compressible Navier-Stokes equation around parallel flows
- MI2010-34 Nobushige KUROKAWA, Masato WAKAYAMA & Yoshinori YAMASAKI
Milnor-Selberg zeta functions and zeta regularizations
- MI2010-35 Kissani PERERA & Yoshihiro MIZOGUCHI
Laplacian energy of directed graphs and minimizing maximum outdegree algorithms
- MI2010-36 Takanori YASUDA
CAP representations of inner forms of $Sp(4)$ with respect to Klingen parabolic subgroup
- MI2010-37 Chikashi ARITA & Andreas SCHADSCHNEIDER
Dynamical analysis of the exclusive queueing process
- MI2011-1 Yasuhide Fukumoto & Alexander B. Samokhin
Singular electromagnetic modes in an anisotropic medium
- MI2011-2 Hiroki Kondo, Shingo Saito & Setsuo Taniguchi
Asymptotic tail dependence of the normal copula
- MI2011-3 Takehiro Hirotsu, Hiroki Kondo, Shingo Saito, Takuya Sato, Tatsushi Tanaka & Setsuo Taniguchi
Anderson-Darling test and the Malliavin calculus
- MI2011-4 Hiroshi Inoue, Shohei Tateishi & Sadanori Konishi
Nonlinear regression modeling via Compressed Sensing
- MI2011-5 Hiroshi Inoue
Implications in Compressed Sensing and the Restricted Isometry Property
- MI2011-6 Daeju Kim & Sadanori Konishi
Predictive information criterion for nonlinear regression model based on basis expansion methods
- MI2011-7 Shohei Tateishi, Chiaki Kinjo & Sadanori Konishi
Group variable selection via relevance vector machine