Variable and boundary selection for functional data via multiclass logistic regression modeling

Hidetoshi Matsui

Faculty of Mathematics, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
hmatsui.math.kyushu-u.ac.jp

Abstract: $\ell_1$ penalties such as the lasso provide solutions with some coefficients to be exactly zeros, which lead to variable selection in regression settings. They also can select variables which affect the classification by being applied to the logistic regression model. We focus on the form of $\ell_1$ penalties in logistic regression models for functional data, especially in the case for classifying the functions into three or more groups. We provide penalties that appropriately select variables in the functional multinomial regression modeling. Simulation and real data analysis show that we should select the form of the penalty in accordance with the purpose of the analysis.

Key Words and Phrases: Functional Data Analysis, Lasso, Logistic regression model, Model selection, Regularization.

1 Introduction

Variable selection is one of crucial issues in regression analysis. The lasso by Tibshirani (1996) and its offshoots provide us an unified approach for estimation and variable selection problems, and therefore they are broadly applied in several fields (see, e.g. Hastie et al., 2009). We can also select variables which affect to classification problems by applying these types of penalties to logistic regression models (Krishnapuram et al., 2005; Park and Hastie, 2007; Friedman et al., 2010). The logistic regression model is one of the most useful tools for classifying data by providing posterior probabilities which group the data belong to.

In this paper we consider the cases for classifying data into three or more groups using the multiclass logistic regression model, which contains multiple parameters in each variable, as well as a multivariate linear model. Several works have been provided for multivariate linear or multiclass logistic regression models. Turlach et al. (2005) proposed a new penalty to estimate multivariate linear models. They imposed an $\ell_1$ sum of maximum absolute values ($\ell_\infty$ norm) of coefficients with respect to multiple responses, and also generalized it to the $\ell_1$ sum of $\ell_q$ ($q \geq 1$) penalties. Then afterwards Yuan et al. (2007) and Obozinski et al. (2011) let the penalty denote $\ell_1/\ell_q$ and investigated theoretical properties of it. Furthermore, Obozinski et al. (2010) proposed a new algorithm for estimating the multinomial logistic regression model with the $\ell_1/\ell_q$ regularization for $q = 1, 2$. 
When the data to be classified are repeatedly measured over time, it is natural that they are represented by some functional forms. Ramsay and Silverman (2005) established this type of analysis and call it the Functional data analysis (FDA). FDA is one of the most useful methods for effectively analyzing such discretely observed data and has received considerable attentions in various fields (Ramsay and Silverman, 2002; Ferraty and Vieu, 2006; Ferraty, 2011). The basic idea behind FDA is to express repeated measurement data as a smooth function and then draw information from the collection of functional data. FDA includes extensions of traditional analyses such as principal component, discriminant and regression analysis (James et al., 2000; James, 2002).

In this paper we consider the variable selection problem for classifying functional data using the logistic regression model via the sparse regularization. Repeated measurement data are represented by basis expansions, and then the functional logistic regression model is estimated by the penalized maximum likelihood method with the help of sparsity penalties. We propose a new penalty, denoted by $\ell_1/\ell_q$ with $q = 1, 2$ here, for appropriately estimating and selecting variables or boundaries for the multiclass functional logistic regression model. Furthermore, the estimated model is evaluated by the model selection criterion since the model evaluation problem is a crucial issue. We examine the proposed method through Monte Carlo simulations and real data examples.

This paper is organized as follows. Section 2 provides a multiclass logistic regression model for functional data. Section 3 shows the method for estimating and evaluating the model. We apply the proposed method to the analysis of simulated data and real data in Section 4 and Section 5 respectively. Concluding remarks are given in Section 6.

2 Multinomial logistic regression model for functional data

Suppose that we have $n$ sets of functional data and a class label $\{ (x_\alpha(t), g_\alpha); \alpha = 1, \ldots, n \}$, where $x_\alpha(t) = (x_{\alpha 1}(t), \ldots, x_{\alpha p}(t))^T$ are predictors given as functions and $g_\alpha \in \{1, \ldots, L\}$ are classes which $x_\alpha$ belongs to. In the classification setting, we apply the Bayes rule which assigns $x_\alpha$ to class $g_\alpha = l$ with the maximum posterior probability given $x_\alpha$, denoted by $Pr(g_\alpha = l|x_\alpha)$. Then the logistic regression model is given by the log-odds of posterior probabilities:

$$\log \left\{ \frac{Pr(g_\alpha = l|x_\alpha)}{Pr(g_\alpha = L|x_\alpha)} \right\} = \beta_{l0} + \sum_{j=1}^{p} \int x_{\alpha j}(t)\beta_{lj}(t)dt,$$

where $\beta_{l0}$ is an intercept and $\beta_{lj}(t)$ are coefficient functions. We consider that $x_{\alpha j}(t)$ is expressed by basis expansions by

$$x_{\alpha j}(t) = \sum_{m=1}^{M_j} w_{\alpha jm} \phi_{jm}(t) = w_{\alpha j}^T \phi_j(t),$$

2
where $\phi_j(t) = (\phi_{j1}(t), \ldots, \phi_{jM_j}(t))^T$ are vectors of basis functions such as $B$-splines or radial basis functions, $w_{\omega_j} = (w_{\omega_j1}, \ldots, w_{\omega_jM_j})$ are coefficient vectors. Since the data are originally observed at discrete time points, we obtain functional data $x_{\omega_j}(t)$ using a smoothing method with basis expansions in advance. In other words, $w_{\omega_j}$ are obtained before constructing the functional logistic regression model (1). Details of the smoothing method are described in Araki et al. (2009b). Furthermore, $\beta_{lj}(t)$ are also expressed by basis expansions

$$
\beta_{lj}(t) = \sum_{m=1}^{M_j} b_{ljm} \phi_{jm}(t) = b_{lj}^T \phi_j(t),
$$

where $b_{lj} = (b_{lj1}, \ldots, b_{ljM_j})^T$ are vectors of coefficient parameters.

Using a notation $\pi_l(x_\alpha; \theta) = \Pr(g_\omega = l | x_\alpha)$ with $\theta = (b_1^T, \ldots, b_{(L-1)}^T)^T$ and $b_l = (\beta_{l0}, b_{l1}^T, \ldots, b_{lp}^T)^T$ since it is controlled by $\theta$, we can re-express the functional logistic regression model (1) as

$$
\log \left\{ \frac{\pi_l(x_\alpha; \theta)}{\pi_L(x_\alpha; \theta)} \right\} = \beta_{l0} + \sum_{j=1}^{p} w_{\alpha_j}^T \Phi_j b_{lj} = z_\alpha^T \beta_l,
$$

where $z_\alpha = (1, w_{\alpha_1}^T \Phi_1, \ldots, w_{\alpha_p}^T \Phi_p)$ and $\Phi_j = \int \phi_j^T(t) \phi_j(t)$. It follows from (1) that the posterior probability is given by

$$
\pi_l(x_\alpha; \theta) = \frac{\exp (z_\alpha^T b_l)}{1 + \sum_{h=1}^{L-1} \exp (z_\alpha^T b_h)} \quad (l = 1, \ldots, L - 1),
$$

$$
\pi_L(x_\alpha; \theta) = \frac{1}{1 + \sum_{h=1}^{L-1} \exp (z_\alpha^T b_h)}.
$$

We define vectors of response variables $y_\alpha$ which indicate class labels as

$$
y_\alpha = (y_{\alpha 1}, \ldots, y_{\alpha(L-1)})^T = \begin{cases} 
(0, \ldots, 0, l, 0, \ldots, 0)^T & \text{if } g_\omega = l, \quad l = 1, \ldots, L - 1, \\
(0, \ldots, 0)^T & \text{if } g_\omega = L.
\end{cases}
$$

Then the functional logistic regression model has the following probability function

$$
f(y_\alpha | x_\alpha; \theta) = \prod_{l=1}^{L-1} \pi_l(x_\alpha; \theta)^{y_{\alpha l}} \pi_L(x_\alpha; \theta)^{1-\sum_{h=1}^{L-1} y_{\alpha h}}.
$$

3 Estimation by the sparse regularization

In order to construct the statistical model, we estimate the functional logistic model (1) by the penalized likelihood method, which maximizes the penalized log-likelihood function given in the form of

$$
l_\lambda(\theta) = l(\theta) - n \lambda P(\theta),
$$

where $l(\theta)$ is the log-likelihood function and $P(\theta)$ is a penalty function.
respectively, and therefore they are regarded as natural extensions of the ℓ₁ type penalties on \( P(\cdot) \). Here we adopt the lasso as the ℓ₁ type penalty in this paper. Now we consider two types of penalties for \( P(\cdot) \):

\[
P_{(1)}(b) = \sqrt{M_j} \sum_{j=1}^{p} \sum_{l=1}^{L-1} \| b_{lj} \|_2, \quad P_{(2)}(b) = \sqrt{M_j(L-1)} \sum_{j=1}^{p} \left\{ \sum_{l=1}^{L-1} \| b_{lj} \|_2^2 \right\}^{\frac{1}{2}}.
\]

(6)

In order to select variables appropriately we need to treat \( w_{\alpha j_1}, \ldots, w_{\alpha j_M} \) as grouped variables and thus apply the group lasso (Yuan and Lin, 2006) to the corresponding coefficients \( b_{lj_1}, \ldots, b_{lj_M} \), otherwise we fail to select functional variables (Matsui and Konishi, 2011). In addition, \( P_{(1)} \) imposes ℓ₁ norms of coefficient parameters for each class. On the other hand, \( P_{(2)} \) treats \( L - 1 \) parameters as grouped parameters gain. The former penalty shrinks each coefficient towards exactly zero, whereas the latter shrinks all the \( L - 1 \) parameter vectors in the \( j \)-th variable towards zero simultaneously. We denote \( P_{(1)} \) and \( P_{(2)} \) as \( \ell_1 \ell_2 / \ell_1 \) and \( \ell_1 \ell_2 / \ell_2 \) penalties respectively. If we consider the case \( M_j = 1 \) for all \( j \), they correspond to \( \ell_1 / \ell_q \) penalties by Obozinski et al. (2010, 2011) with \( q = 1, 2 \) respectively, and therefore they are regarded as natural extensions of the \( \ell_1 / \ell_q \) penalties.

We can also express the \( \ell_1 \ell_2 / \ell_q \) norm in the following form.

\[
P_{(q)}(b) = C \sum_{j=1}^{p} \left\{ \sum_{l=1}^{L-1} \| b_{lj} \|_2 \right\}^{\frac{1}{q}},
\]

here we denoted a constant independent of \( b \) as \( C \).

Since the penalized log-likelihood function involves the ℓ₁ norm of coefficients it is difficult to derive estimates analytically. We apply the local quadratic approximation to the penalty (Tibshirani, 1996; Fan and Li, 2001). Then parameters are updated in the following form:

\[
\hat{b}^{(k+1)} = \hat{b}^{(k)} - \left\{ \frac{\partial^2 l(\hat{b})}{\partial \hat{b} \partial \hat{b}^T} \bigg|_{\hat{b}^{(k)}} - n \Sigma(\hat{b}^{(k)}) \right\}^{-1} \left\{ \frac{\partial l(\hat{b})}{\partial \hat{b}} \bigg|_{\hat{b}^{(k)}} - n \Sigma(\hat{b}^{(k)}) \hat{b}^{(k)} \right\},
\]

where \( \Sigma(\hat{b}) \) is given by

\[
\Sigma(\hat{b}) = \text{diag} \left\{ \Sigma(\hat{b}_1), \ldots, \Sigma(\hat{b}_{(L-1)}) \right\},
\]

\[
\Sigma(\hat{b}_l) = \begin{cases} 
\text{diag} \left\{ \frac{P_{(1)}(|\hat{b}_{lj_1}|)}{|\hat{b}_{lj_1}|}, \ldots, \frac{P_{(1)}(|\hat{b}_{lj_p}|)}{|\hat{b}_{lj_p}|} \right\} & \text{for } \hat{P}_{(1)}, \\
\text{diag} \left\{ \frac{P_{(2)}(|\hat{b}_{lj_1}|)}{|\hat{b}_{lj_1}|}, \ldots, \frac{P_{(2)}(|\hat{b}_{lj_p}|)}{|\hat{b}_{lj_p}|} \right\} & \text{for } \hat{P}_{(2)}
\end{cases}
\]

(\( \hat{b}_{lj} = (b_{lj_1}, \ldots, b_{lj_{(L-1)}})^T \)).

This update is continued until convergence, and then we obtain an estimated coefficient vector \( \hat{b} \).
Statistical models estimated by the penalized likelihood method depend on regularization parameters and it is a crucial issue to select appropriate values of them. In this case we need to select an appropriate value of \( \lambda \) in the log-likelihood function (5). We use a model selection criterion BIC originally proposed by Schwarz (1978). For the sparse regularization problem, Wang et al. (2007) proved that the BIC select the true model consistently for the SCAD regularization setting. The BIC for evaluating models estimated by the penalized maximum likelihood method with the \( \ell_1 \) penalty is given by

\[
\text{BIC} = -2l(\hat{b}) + \tilde{df} \log n
\]

\[
= \sum_{\alpha=1}^{n} \sum_{l=1}^{L-1} \left( y_{al} \log \pi_l(z_\alpha; \hat{b}) + \left( 1 - \sum_{h=1}^{L-1} y_{ah} \right) \log \pi_L(z_\alpha; \hat{b}) \right) + \tilde{df} \log n,
\]

where \( \tilde{df} \) is an effective degrees of freedom. Although Zou et al. (2007) derived a degrees of freedom for the model estimated by the lasso-type regularization, it is not given in the logistic settings. Konishi et al. (2004) proposed a model selection criterion GBIC for evaluating models estimated by the regularization method, though, it needs a second derivative of the penalized likelihood function. On the other hand, they also derived AIC or BIC type criterion with effective degrees of freedoms in the logistic regression model. Using this idea, the degrees of freedom here is given by

\[
\tilde{df} = \text{tr}\{W\tilde{Z}_A(Z_A^TW\tilde{Z}_A + n\Sigma(\hat{b}_A))^{-1}\tilde{Z}_A^T\},
\]

where

\[
\tilde{Z} = I_{L-1} \otimes Z, \quad Z = (z_1^T, \ldots, z_n^T),
\]

\[
W^{(k)} = \begin{pmatrix} W_{hl}^{(k)} \end{pmatrix}_{hl},
\]

\[
W_{hl}^{(k)} = \begin{cases} \text{diag}\{\pi_l(z_1; \hat{b})(1 - \pi_l(z_1; \hat{b})), \ldots, \pi_l(z_n; \hat{b})(1 - \pi_l(z_n; \hat{b}))\} & (h = l), \\ \text{diag}\{-\pi_h(z_1; \hat{b})\pi_l(z_1; \hat{b}), \ldots, -\pi_h(z_n; \hat{b})\pi_l(z_n; \hat{b})\} & (h \neq l) \end{cases}
\]

and the suffix \( A \) denotes an active set of \( b \). We select the \( \lambda \) which minimizes the BIC and then adopt the corresponding model as an optimal model.

### 4 Simulation

In order to investigate the effectiveness and the behavior of the proposed method, we conducted Monte Carlo simulations. We simulated multiple predictors, essentially given as functions having three classes. In this example we aim to examine whether the proposed method appropriately select variables which affect to the classification.
First, we generated $2n$ sets of $p$ predictors each of which were repeatedly measured at several time points; $\{x_{ai1}, \ldots, x_{aip}\}$, where $x_{aij} (j = 1, \ldots, p)$ are assumed to be obtained from $x_{aij} = u_{aij}(t_i) + \varepsilon_{aij}$ with uniformly distributed $n_i$ observational points $t_i$ and $\varepsilon_{aij} \sim N(0, (\sigma R_{aij})^2)$, $R_{aij} = \max_i (u_{aij}(t_i)) - \min_i (u_{aij}(t_i))$.

We divided the data with sample size $2n$ in half in order to use them as training and test data. Here we consider a three-class classification problem, and therefore we trisected the data and then assigned each class with them. We set true functions $u_{aij}(t)$ as follows:

$$u_{a1}(t) = t^3 + a_1 t^2 + a_2 t + a_3, \quad u_{a2}(t) = b_1 \sin(b_2 \pi (t - b_3)), \quad u_{a3}(t) = c_1 \exp(c_2 t)(t - c_3)^2, \quad u_{a4}(t) = d_1 \sin(\pi t) + d_2, \quad t \in [-1, 1],$$

where they are controlled by following random numbers for classes $g = 1, 2, 3$ respectively.

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$ $\sim$ $N(3,0,0.1^2)$, $N(2,0,0.1^2)$, $N(4,0,0.1^2)$,</td>
<td>$u_1 : a_2$ $\sim$ $N(0,5,0.2^2)$, $N(0,3,0.2^2)$, $N(0,0,2^2)$,</td>
<td>$a_3$ $\sim$ $N(1,0,0.5^2)$, $N(1,0,0.5^2)$, $N(2,0,0.5^2)$,</td>
</tr>
<tr>
<td>$b_1$ $\sim$ $U(0,4,0.7)$, $U(0,8,1.1)$, $U(0,4,0.7)$,</td>
<td>$u_2 : b_2$ $\sim$ $U(0,9,1.2)$, $U(0,4,0.7)$, $U(0,9,0.7)$,</td>
<td>$b_3$ $\sim$ $N(0,2,0.1^2)$, $N(0,5,0.1^2)$, $N(0,2,0.7^2)$,</td>
</tr>
<tr>
<td>$c_1$ $\sim$ $U(0,7,1.1)$, $U(0,6,0.9)$, $U(0,6,0.9)$,</td>
<td>$u_3 : c_2$ $\sim$ $U(0,7,1.2)$, $U(0,8,1.4)$, $U(0,8,1.4)$,</td>
<td>$c_3$ $\sim$ $N(0,8,1.4)$, $N(0,5,0.1^2)$, $N(0,5,0.1^2)$,</td>
</tr>
<tr>
<td>$u_4 : d_1$ $\sim$ $U(0,5,3.0)$, $U(0,5,3.0)$, $U(0,5,3.0)$,</td>
<td>$d_2$ $\sim$ $N(0,2,0.5)$, $N(0,2,0.5)$, $N(0,2,0.5)$,</td>
<td></td>
</tr>
</tbody>
</table>

Note that there are no classification boundaries for Class 1 and 3 on $u_2$, Class 2 and 3 on $u_3$ and all classes on $u_4$, as seen in Table 1 since they respectively have same settings for the random number generation. Especially $u_4$ does not affect the classification. These facts show that the coefficient functions in the model (1) should be $\beta_{12}(t) = 0$, $\beta_{23}(t) = 0$ and $\beta_{14}(t) = \beta_{24}(t) = 0$.

As the first step of the analysis, we converted longitudinal predictors $x_{aij}$ into functions. Since they contain additive noises we applied the smoothing method with basis expansions, then obtained functional data sets $x_{aij}(t)$. We used radial basis functions with the idea of $B$-splines by Kawano and Konishi (2007) for basis functions $\phi_j(t)$ in (2) and (3). In order to reduce computational burden, numbers of basis functions are supposed to be 6 for all variables. Then after constructing the functional logistic regression model, we estimated it by the maximum penalized likelihood method with penalties $P_{(1)}$ and $P_{(2)}$ in (6) and evaluated the model by BIC. We repeated them 100 times for $n = 75, 150, 300$ and $\sigma = 0.05, 0.2$ and then investigated about results of errors and variable selections for two penalties.

Table 1 shows numbers selected by the above method with penalties $P_{(1)}$ and $P_{(2)}$ for 100 repetition, where “$\beta_{ij}\beta_{kj}$” in the table denotes numbers which both $\beta_{ij}$ and $\beta_{kj}$...
are estimated to be non-zeros, ”$\beta_{lj}$” for $l = 1, 2$ denotes numbers which either $\beta_{1j}$ or $\beta_{2j}$ is estimated to be non-zero respectively and ”$\phi$” is numbers which both coefficients are estimated to be zeros. We can find that the penalty $P(1)$ shrinks some of coefficient functions to be zeros appropriately, especially for variable 1 and 3. Thus it selects which classification boundary is important or not for each variable. On the other hand, $P(2)$ selects all or no coefficients for each variable owing to the form of the penalty and hence it cannot select boundaries individually like $P(1)$. However, it excludes variable 4 completely more than $P(1)$, which indicates that $P(2)$ is more suitable for the purpose of only excluding variables themselves which affect the classification. We also obtained training and test errors and selected values of regularization parameters in Table 2. It indicates that the penalty $P(1)$ gives smaller test errors than $P(2)$ in all cases.

Note that we did not mention a classification boundary between Class 1 and 2. We can directly derive the coefficients for the boundary as $\hat{\beta}_{1j} - \hat{\beta}_{2j}$, and if there are no boundaries for Class 1 and 2 there should be $\hat{\beta}_{1j} = \hat{\beta}_{2j}$. However, the above method does not estimate like it and therefore we cannot estimate appropriately, which reminds to be
5 Real data example

We applied the functional logistic regression modeling to the analysis of yeast cell cycle gene expression data. Spellman et al. (1998) measured expression profiles for 6,178 genome-wide genes in the yeast genome using cDNA microarrays over about two cell cycles. The data contain 77 microarrays consisting of several types of time course synchronization; "cln3" (2 points), "clb2" (2 points), "α-factor" (18 points), "cdc15" (24 points), "cdc28" (17 points) and "elu" (14 points). Spellman et al. (1998) classified 800 genes into 5 groups, G1, G2/M, M/G1, S and S/G2, by the clustering method from the above 77 experiments. Figure 2 shows examples for each synchronization. Araki et al. (2009a) classified genes by using the "cdc15" experiments as functional data and examined the misclassified data from the posterior probabilities. Here we aim to confirm that these experiments actually affect the classification.

Since there are some missing values in expression profiles we excluded some genes from data by following two rules: (1) Genes with at least one missing value for "cln3" and "clb2" respectively are excluded. (2) Those with more than 10 missing values in all for
"α-factor", "cdc15", "cdc28" and "elu" are also excluded. Although the number of genes with no missing values are only 72, we can easily apply the regression model even if there are some (not excessively many) missing values by converting them into functional data. The resulting 657 genes are used for this analysis. First we converted time-course data except for "cln3" and "clb2" into functions. They are expressed by basis expansions with 4 basis functions, which was selected in the functionalization step. Remaining variables "cln3" and "clb2", each of which have only 2 time points, were treated as vector data rather than functional data, and then we treated variables corresponding to 2 time points for each variable as grouped variables. Next we constructed a functional logistic model

\[
\log \left\{ \frac{\Pr(g_\alpha = l | x_\alpha)}{\Pr(g_\alpha = L | x_\alpha)} \right\} = \beta_{l0} + \sum_{j=1}^{2} \sum_{j'=1}^{2} x_{\alpha j,j'} \beta_{l j'} + \sum_{j=3}^{6} \int x_{\alpha j}(t) \beta_{lj}(t) dt, \tag{7}
\]

which is a special case of (1), where \( X_j \ (j = 1, \ldots, 6) \) respectively correspond to "cln3", "clb2", "α-factor", "cdc15", "cdc28" and "elu". The model was estimated by the penalized likelihood method and evaluated by BIC. Furthermore we altered the class label \( L \) in the left hand side of (7) and repeatedly estimated in order to investigate all coefficients for classification boundaries. We repeated this process for 100 bootstrap samples, and then investigated which variable or boundary affects the classification.

Table 3 shows the numbers of selected variables in 100 bootstrap samples for the penalty \( P_{(1)} \). We can find that most coefficients are estimated to be non-zeros. However, the coefficient for the boundary between M/G1 and S/G2 is hardly selected for "clb2", and similarly that between M/G1 and S is often excluded. It reveals that the variable "clb2" does not affect the classification. On the other hand, the penalty \( P_{(2)} \) selected all variables for 100 repetitions. It indicates the variables themselves are actually relevant to the classification.

6 Concluding remarks

We have proposed a form of penalty for constructing functional multinomial logistic regression models. We derived the estimation and evaluation procedures for the model with the \( \ell_1 \ell_2/\ell_q \) penalty for \( q = 1, 2 \). The model was fitted by the penalized maximum likelihood method, and the regularization parameter involved in the model was selected by the model selection criterion. Monte Carlo simulations were conducted in order to investigate the effects for the accuracy of prediction and variable selection. Results show that \( \ell_1 \ell_2/\ell_1 \) and \( \ell_1 \ell_2/\ell_2 \) penalties give different result and therefore we should select these penalties for different uses.

As described in the end of Section 4, there are cases that this method does not estimate coefficients appropriately at one time. It will be a topic for future research. Furthermore we will focus on the theoretical investigation for the functional regression models with the sparse regularization.
References


— (2005), *Functional data analysis*, Wiley Online Library.


Table 1: Numbers of selected variables. Top:$P(1)$, bottom:$P(2)$.

<table>
<thead>
<tr>
<th>$P(1)$</th>
<th>$\sigma$</th>
<th>0.05</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$j$</td>
<td>$\beta_{1j}\beta_{2j}$</td>
<td>$\beta_{1j}$</td>
</tr>
<tr>
<td>75</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37 0 63 0</td>
<td>33 0 67 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0 100 0 0</td>
<td>1 99 0 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0 25 6 69</td>
<td>12 45 16 27</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>32 0 68 0</td>
<td>49 0 51 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0 100 0 0</td>
<td>1 99 0 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1 33 2 64</td>
<td>15 42 14 29</td>
</tr>
<tr>
<td>300</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>45 0 55 0</td>
<td>34 0 66 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0 100 0 0</td>
<td>1 99 0 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2 32 5 61</td>
<td>26 40 15 19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$P(2)$</th>
<th>$\sigma$</th>
<th>0.05</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>$j$</td>
<td>$\beta_{1j}\beta_{2j}$</td>
<td>$\beta_{1j}$</td>
</tr>
<tr>
<td>75</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>98 0 0 2</td>
<td>91 0 0 9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100 0 0 0</td>
<td>99 0 0 1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>18 0 0 82</td>
<td>54 0 0 46</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>22 0 0 78</td>
<td>51 0 0 49</td>
</tr>
<tr>
<td>300</td>
<td>1</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>100 0 0 0</td>
<td>100 0 0 0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>26 0 0 74</td>
<td>71 0 0 29</td>
</tr>
</tbody>
</table>
Table 2: Averages of 100 training, test errors (%) and the regularization parameter $\lambda$ (and its standard deviations in parenthesis). The top is for $P_{(1)}$ and the bottom is for $P_{(2)}$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.05</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>75</td>
<td>0.04</td>
<td>0.41</td>
</tr>
<tr>
<td>150</td>
<td>0.10</td>
<td>0.33</td>
</tr>
<tr>
<td>300</td>
<td>0.10</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.05</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>75</td>
<td>0.08</td>
<td>0.80</td>
</tr>
<tr>
<td>150</td>
<td>0.15</td>
<td>0.45</td>
</tr>
<tr>
<td>300</td>
<td>0.12</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Table 3: Numbers of selected variables for $P_{(1)}$. Signs $i/j$ below indicate classification boundaries between $i$ and $j$ for $i, j = 1, \ldots, 5$ with 1:G1, 2:G2/M, 3:M/G1, 4:S, 5:S/G2.

<table>
<thead>
<tr>
<th></th>
<th>cln3</th>
<th>clb2</th>
<th>$\alpha$</th>
<th>cdc15</th>
<th>cdc28</th>
<th>elu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>1/3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1/4</td>
<td>94</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2/3</td>
<td>100</td>
<td>76</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2/4</td>
<td>100</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>2/5</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3/4</td>
<td>100</td>
<td>47</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3/5</td>
<td>100</td>
<td>16</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4/5</td>
<td>100</td>
<td>82</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>