Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

www.czasopisma.uni.lodz.pl/foe/
2(334) 2018

DOI: http://dx.doi.org/10.18778/0208-6018.334.04

Mirosław Krzyśko
The President Stanisław Wojciechowski State University of Applied Sciences in Kalisz, Interfaculty Institute of Mathematics and Statistics, mkrzysko@amu.edu.pl

Łukasz Smaga
Adam Mickiewicz University in Poznań, Faculty of Mathematics and Computer Science, ls@amu.edu.pl

Selected Robust Logistic Regression Specification for Classification of Multi-dimensional Functional Data in Presence of Outlier

Abstract: In this paper, the binary classification problem of multi-dimensional functional data is considered. To solve this problem a regression technique based on functional logistic regression model is used. This model is re-expressed as a particular logistic regression model by using the basis expansions of functional coefficients and explanatory variables. Based on re-expressed model, a classification rule is proposed. To handle with outlying observations, robust methods of estimation of unknown parameters are also considered. Numerical experiments suggest that the proposed methods may behave satisfactory in practice.

Keywords: basis functions representation, classification problem, functional regression analysis, logistic regression model, multi-dimensional functional data, robust estimation

JEL: C38, C13

1. Introduction

From the end of the 1990s, Functional Data Analysis (FDA) has become increasingly popular and is now one of the major research fields in statistics. In FDA, the theory and practice of statistical methods are studied in situations where the available data are functions. Such data appear and are analized in different fields of applications, including economics (e.g., the GDP per capita, Górecki, Łaźniewska, 2013; the level of income, Jaworski, Pietrzykowski, 2014), meteorology (e.g., the temperatures, pressure, etc., in a given location, Collazos, Dias, Zambom, 2016), and many others (see, for example, Ramsay, Silverman, 2002, where the illustration of certain FDA methods through the study of specific case studies with real data is given). Particular problems of FDA considered in the literature are as follows: analysis of variance (Zhang, 2013; Górecki, Smaga, 2015; 2017), canonical correlation analysis (Krzyśko, Waszak, 2013), classification problem (James, Hastie, 2001; Górecki, Krzyśko, Wołyński, 2015), cluster analysis (Giacofci et al., 2013), nonparametric analysis (Ferraty, Vieu, 2006), outlier detection (Febrero-Bande, Galeano, González-Manteiga, 2007; 2008; Hubert, Rousseeuw, Segaert, 2015), principal component analysis (Ramsay, Silverman, 2005; Kayano, Konishi, 2009; Horváth, Kokoszka, 2012), regression analysis (Chiou, Müller, Wang, 2004; Chiou, Yang, Chen, 2016; Matsui, Konishi 2011; Collazos, Dias, Zambom, 2016).

In this paper, we consider one of the main problems of FDA, namely, the binary classification problem of multi-dimensional functional data. Recently, Górecki, Krzyśko, Wołyński (2015) studied this problem by using multivariate functional regression techniques, e.g., functional logistic regression model, which performed best on real data examples. It is worth noting that for estimation of unknown parameters, they used standard maximum likelihood estimation method. However, in the presence of outliers, this method may behave poorly, which probably adversely affects the classification process. In this article, we propose an extension of the method of Górecki, Krzyśko, Wołyński (2015). More precisely, we study a more general representation of the functional logistic regression model and use the robust estimation methods in logistic regression. Numerical results indicate that the new classification rules behave promisingly and may be reasonable competitors to existing methods. The binary classification rules can be extended for multi-label classification problems by using existing techniques (see, for instance, Krzyśko, Wołyński, 2009).

The rest of the present paper is organized as follows. In Section 2, we introduce a binary classification problem for multivariate functional data. Section 3 contains the construction of the functional logistic regression model and its re-expression based on the basis functions representation of coefficients and explanatory variables. The classification rule based on this model is also presented there. In Section 4, we review robust estimation methods in logistic regression model.

The accuracy of the proposed methods and their comparison with existing ones is demonstrated using two real functional data sets in Section 5. Section 6 concludes the article.

2. Binary classification problem for functional data

In binary classification problem, we have to determine a procedure assigning a given object to one of two populations. Classically, the objects are characterized by p scalar features, and then the observations are p-dimensional random vectors. In functional data analysis, the features are given by functions observed at possibly different time points. Below, we precisely formulate the problem of binary classification for multi-dimensional functional data.

Assume that we have the learning sample $\{(\mathbf{x}_i(t), Y_i): i = 1, \ldots, N\}$, where $\mathbf{x}_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))'$ are p-dimensional vectors of random functions describing the objects, and $Y_i \in \{0, 1\}$ are the labels of classes to which the objects belong. The functions $x_{ij}(t)$, $i = 1, \ldots, N$ are supposed to belong to the Hilbert space of square integrable functions over $T_j = [a_j, b_j]$, $a_j, b_j \in R$, $j = 1, \ldots, p$. This space will be denoted by $L_2(T_j)$, $j = 1, \ldots, p$.

To solve the classification problem described above, the Bayes rule can be used. This rule assigns $\mathbf{x}$ to class $Y = k$ with the maximum posterior probability given $\mathbf{x}$, denoted by $P(Y = k|\mathbf{x})$, $k = 0, 1$. The Bayesian classifier then takes the form (see, for example, Krzyśko et al., 2008):

$$d(\mathbf{x}) = \begin{cases} 1, & P(Y = 1|\mathbf{x}) \geq P(Y = 0|\mathbf{x}), \\ 0, & P(Y = 1|\mathbf{x}) < P(Y = 0|\mathbf{x}). \end{cases} \tag{1}$$

In classical classification problem, it is well known that $P(Y = 1|\mathbf{x}) = E(Y|\mathbf{x}) = r(\mathbf{x})$, where $r(\mathbf{x})$ is the regression function of the random variable Y with respect to the random vector $\mathbf{x}$. Then the classifier (1) can be rewritten as follows:

$$d(\mathbf{x}) = \begin{cases} 1, & r(\mathbf{x}) \geq 1/2, \\ 0, & r(\mathbf{x}) < 1/2. \end{cases} \tag{2}$$

Different regression functions as well as their estimates are able to be used in classifier (2), e.g., linear or logistic regression function.

Górecki, Krzyśko, Wołyński (2015) applied the above idea to the classification problem for multi-dimensional functional data. They used four functional regression methods as $r(\mathbf{x})$, i.e., multivariate linear regression, logistic regression, local linear regression smoothers and Nadaraya-Watson kernel estimation method. The best numerical results were obtained by applying the functional logistic regression model. Therefore, we consider this method and propose possible improvement of it in the next Sections.

3. Functional logistic regression model

In this Section, we present the functional logistic regression model in more general form, as in Górecki, Krzyśko, Wołyński (2015). Using this model, we then propose the classification rule for functional data.

Let us introduce the functional logistic regression model by using the assumptions and notation of Section 2. The variables Y_i, $i = 1, \ldots, N$ are assumed to be independent Bernoulli response variables. The components of the vector $\mathbf{x}_i(t)$ are considered as explanatory functional variables. Let observations follow the functional logistic regression model of the form:

$$P(Y_i = 1|\mathbf{x}_i) = \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \int_{T_j} x_{ij}(t)\beta_j(t)dt\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^{p} \int_{T_j} x_{ij}(t)\beta_j(t)dt\right)}, \tag{3}$$

where β_0 is the intercept and $\beta_j(t) \in L_2(T_j)$, $j = 1, \ldots, p$ are the unknown coefficient functions.

The model (3) can be rewritten by using the basis functions representation as described below. Since $x_{ij}(t), \beta_j(t) \in L_2(T_j)$, $j = 1, \ldots, p$, these functions can be approximated arbitrarily well by a linear combination of a sufficiently large number of basis functions $\{\varphi_{jm}\}_{m=1}^{\infty}$ of $L_2(T_j)$ (Ramsay, Silverman, 2005). Thus, assume that the functions $x_{ij}(t)$ and $\beta_j(t)$ can be represented as follows:

$$\begin{gathered} x_{ij}(t) = \sum_{m=1}^{B_j} w_{ijm}\varphi_{jm}(t) = \mathbf{w}_{ij}'\boldsymbol{\varphi}_j(t), \\ \beta_j(t) = \sum_{m=1}^{B_j} b_{jm}\varphi_{jm}(t) = \mathbf{b}_j'\boldsymbol{\varphi}_j(t), t \in T_j, \end{gathered} \tag{4}$$

where

$$i = 1, \ldots, N, j = 1, \ldots, p, \mathbf{w}_{ij}' = \left(w_{ij1}, \ldots, w_{ijB_j}\right) \text{ and } \mathbf{b}_j' = \left(b_{j1}, \ldots, b_{jB_j}\right)$$

are the vectors of unknown coefficients, and

$$\boldsymbol{\varphi}_j'(t) = \left(\varphi_{j1}(t), \ldots, \varphi_{jB_j}(t)\right)$$

are the vectors of basis functions. For each $j = 1, \ldots, p$, the vectors $\mathbf{w}_{ij}$ can be estimated by using the functional observations $x_{ij}(t)$, $i = 1, \ldots, N$ and the least squares method (see Krzyśko, Waszak, 2013). The truncation parameters B_j and the basis functions φ_{jm} may be chosen in such a way to improve the solution of the problem under consideration, e.g., reduce the classification error of a particular classifier. By (4), the model (3) can be re-expressed as follows:

$$P(Y_i = 1|\mathbf{x}_i) = \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \int_{T_j} \mathbf{w}'_{ij} \boldsymbol{\varphi}_j(t) \boldsymbol{\varphi}'_j(t) \mathbf{b}_j dt\right)}{1+\exp\left(\beta_0 + \sum_{j=1}^{p} \int_{T_j} \mathbf{w}'_{ij} \boldsymbol{\varphi}_j(t) \boldsymbol{\varphi}'_j(t) \mathbf{b}_j dt\right)}$$

$$= \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \mathbf{w}'_{ij} \int_{T_j} \boldsymbol{\varphi}_j(t) \boldsymbol{\varphi}'_j(t) dt \mathbf{b}_j\right)}{1+\exp\left(\beta_0 + \sum_{j=1}^{p} \mathbf{w}'_{ij} \int_{T_j} \boldsymbol{\varphi}_j(t) \boldsymbol{\varphi}'_j(t) dt \mathbf{b}_j\right)}, \tag{5}$$

$$= \frac{\exp\left(\beta_0 + \sum_{j=1}^{p} \mathbf{w}'_{ij} \mathbf{J}_{\boldsymbol{\varphi}_j} \mathbf{b}_j\right)}{1+\exp\left(\beta_0 + \sum_{j=1}^{p} \mathbf{w}'_{ij} \mathbf{J}_{\boldsymbol{\varphi}_j} \mathbf{b}_j\right)}$$

$$= \frac{\exp(\beta_0 + \mathbf{w}'_i \mathbf{b})}{1+\exp(\beta_0 + \mathbf{w}'_i \mathbf{b})},$$

where $i = 1, \ldots, N$, $\mathbf{J}_{\boldsymbol{\varphi}_j} := \int_{T_j} \boldsymbol{\varphi}_j(t) \boldsymbol{\varphi}'_j(t) dt$, $j = 1, \ldots, p$, $\mathbf{w}'_i = (\mathbf{w}'_{i1} \mathbf{J}_{\boldsymbol{\varphi}_1}, \ldots, \mathbf{w}'_{ip} \mathbf{J}_{\boldsymbol{\varphi}_p})$ and $\mathbf{b}' = (\mathbf{b}'_1, \ldots, \mathbf{b}'_p)$. The matrix $\mathbf{J}_{\varphi_j}$ is the $B_j \times B_j$ cross product matrix corresponding to basis $\{\varphi_{jm}\}_{m=1}^{\infty}$, $j = 1, \ldots, p$. For an orthonormal basis (e.g., Fourier basis), this matrix is the identity matrix.[1]

Thus, we re-expressed the functional logistic regression model (3) as the logistic regression model (5), where $(\beta_0, \mathbf{b}')'$ is the $(1 + \sum_{j=1}^{p} B_j) \times 1$ vector of unknown parameters. We can use this relationship for estimation problem in functional logistic regression model (3) by using estimation methods for the logistic regression model (5). Let $(\hat{\beta}_0, \hat{\mathbf{b}}')'$ be the estimator of $(\beta_0, \mathbf{b}')'$ obtained under the model (5). Thus we have the following estimator of the regression function:

$$\hat{r}(\mathbf{x}) = \frac{\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})}{1+\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})}. \tag{6}$$

By (2) and (6), we obtain the following classifier for the binary classification problem for functional data presented in Section 2:

$$\hat{d}(\mathbf{x}) = \begin{cases} 1, & \frac{\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})}{1+\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})} \geq 1/2, \\ 0, & \frac{\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})}{1+\exp(\hat{\beta}_0 + \mathbf{w}'\hat{\mathbf{b}})} < 1/2. \end{cases} \tag{7}$$

Now, we have to take into account the estimation problem in the logistic regression model (5). This problem is discussed in the next Section.

[1] For a non-orthonormal basis (e.g., B-spline basis), it can be approximated, for example, by using the function `inprod` from the R package `fda` (Ramsay, Hooker, Graves, 2009; Ramsay et al., 2014; R Core Team, 2017).

4. Robust estimation in logistic regression

For estimating the vector of parameters $\boldsymbol{\gamma} = (\beta_0, \mathbf{b}')'$ in the logistic regression model (5), the maximum likelihood estimator (MLE) is classically used, which is the most efficient estimator (asymptotically). It is defined as

$$\hat{\boldsymbol{\gamma}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\gamma}} \ln L(\boldsymbol{\gamma}; \mathbf{w}_i, Y_i, i = 1, \dots, N) = \\ = \arg\min_{\boldsymbol{\gamma}} \sum_{i=1}^{N} d\big((1, \mathbf{w}_i')\boldsymbol{\gamma};\, Y_i\big), \tag{8}$$

where ln denotes the natural logarithm,

$$\ln L(\boldsymbol{\gamma}; \mathbf{w}_i, Y_i, i = 1, \dots, N) = \sum_{i=1}^{N} l(\boldsymbol{\gamma}; \mathbf{w}_i, Y_i)$$

is the conditional ln-likelihood function,

$$l(\boldsymbol{\gamma}; \mathbf{w}_i, Y_i) = Y_i \ln\left(\frac{\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}{1+\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}\right) + (1 - Y_i)\ln\left(\frac{1}{1+\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}\right)$$

and $d\big((1, \mathbf{w}_i')\boldsymbol{\gamma};\, Y_i\big) = -l(\boldsymbol{\gamma}; \mathbf{w}_i, Y_i)$ is the deviance component. To find $\hat{\boldsymbol{\gamma}}_{\mathrm{MLE}}$, one has to solve the likelihood score equation

$$\sum_{i=1}^{N}\left(Y_i - \frac{\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}{1+\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}\right)(1, \mathbf{w}_i')' = 0. \tag{9}$$

The equations (9) can be solved iteratively by using, for example, the Newton-Raphson method.

Unfortunately, the MLE may behave very poorly in presence of outliers. In functional data, the outliers may also appear, and in particular they may have a negative effect on performance of the classifier (7), when the MLE is used to estimate the parameters of the model (5). To avoid possible drawback of the MLE, it seems to be reasonable to consider robust estimators in classifier (7). For excellent overview and comparison of robust estimation in the logistic regression, we refer to the survey paper by Ahmad, Ramli, Midi (2010). In the remainder of this Section, we describe ideas of robust estimation and certain robust estimators in the logistic regression, which will be compared to the MLE in numerical experiments of Section 5.

The first alternative approach to the MLE is based on weighting the likelihood score function in equation (9), i.e., a robust estimator is the solution of

$$\sum_{i=1}^{N}\left(Y_i - \frac{\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)}{1+\exp\big((1, \mathbf{w}_i')\boldsymbol{\gamma}\big)} - c\big(\mathbf{w}_i, (1, \mathbf{w}_i')\boldsymbol{\gamma}\big)\right)(1, \mathbf{w}_i')'\omega_i = 0,$$

where $c(\mathbf{w}_i, (1, \mathbf{w}_i')\boldsymbol{\gamma})$ is a debiasing factor, i.e., a correction function defined to ensure consistency, and ω_i are the weights depending on $\mathbf{w}_i$, Y_i or both. When the weights depend only on the design, i.e., $\omega_i = \omega(\mathbf{w}_i, (1, \mathbf{w}_i')\boldsymbol{\gamma})$, and $c(\mathbf{w}_i, (1, \mathbf{w}_i')\boldsymbol{\gamma}) = 0$, the estimator obtained in such a way is an MLE computed with weights and is called MALLOWS estimator (Mallows, 1975). For example, the robust Mahalanobis distance of the regressors is a particular weight function ω, i.e., $[(\mathbf{w}_i - \widehat{\mathbf{w}})'\mathbf{S}^{-1}(\mathbf{w}_i - \widehat{\mathbf{w}})]^{1/2}$, where $\widehat{\mathbf{w}}$ and $\mathbf{S}$ are the robust estimators of the center and scatter matrix of the regressors. If the weights depend on the regressors and the response, i.e., $\omega_i = \omega(\mathbf{w}_i, (1, \mathbf{w}_i')\boldsymbol{\gamma}, Y_i)$, the estimators are in the Schweppe class, and they are also known as the conditionally unbiased bounded influence function (CUBIF) estimator (Künsch, Stefanski, Carroll, 1989). In such weights, the differences $Y_i - \exp\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)/\left[1 + \exp\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right]$ are usually used.

The other alternative robust approach is based on modification of the function d in (8). Bianco and Yohai (1996) constructed a consistent (BY) estimator defined by

$$\hat{\boldsymbol{\gamma}}_{\mathrm{BY}} = \arg\min_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \left[\rho\left(d\left((1, \mathbf{w}_i')\boldsymbol{\gamma}; Y_i\right)\right) + \right.$$
$$\left. + G\left(F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right) + G\left(1 - F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right)\right], \tag{10}$$

where

$$\rho(x) = (x - x^2/(2c))I_{(-\infty,c]}(x) + (c/2)I_{(c,\infty)}(x),$$

c is a tuning parameter,

$$G(x) = \int_0^x \rho'(-\ln(u))du, \quad F(x) = \exp(x)/[1 + \exp(x)],$$

and I_A stands for the usual indicator function on the set A ($I_A(x) = 1$ if $x \in A$ and 0 otherwise). In (10),

$$G\left(F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right) + G\left(1 - F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right)$$

is a bias correction term. Bianco and Yohai (1996) also stressed that other choices of the bounded function ρ are possible. To reduce the influence of outliers in the regressor space, Croux and Haesbroeck (2003) proposed to include the weights in (10). The resulting weighted BY (WBY) estimator is given by the formula

$$\hat{\boldsymbol{\gamma}}_{\mathrm{WBY}} = \arg\min_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \omega_i \left[\rho\left(d\left((1, \mathbf{w}_i')\boldsymbol{\gamma}; Y_i\right)\right) + G\left(F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right) + \right.$$
$$\left. + G\left(1 - F\left((1, \mathbf{w}_i')\boldsymbol{\gamma}\right)\right) - G(1)\right],$$

where for instance, $\omega_i = I_{(-\infty,\chi^2_{m,0.975}]}((\mathrm{RMD}_i)^2)$, $\chi^2_{m,0.975}$ is the 0.975-quantile of the central chi-squared distribution with $m = \sum_{j=1}^{p} B_j$ degrees of freedom, and RMD_i is the robust Mahalanobis distance obtained by using the minimum covariance determinant estimator (see Rousseeuw, 1985, and Section 4 of Croux, Haesbroeck, 2003, for description and more details). Since the weights depend solely on the regressors, the WBY estimator remains consistent without any further distribution assumptions. However, the weights used may be too restrictive, resulting in a loss of efficiency of this estimator.

The MLE and four methods of robust estimation in logistic regression model are compared in the next Section in terms of performance of the classifier (7) for functional data.

5. Numerical experiments

In order to test the performance of the classifier (7) described in Section 3, we conducted computational experiments on real functional data sets[2]. The problem of interest is to compare the behavior of this classification rule based on different methods of parameter estimation. More precisely, we consider the MLE and its four robust competitors, i.e., the MALLOWS, CUBIF, BY and WBY estimators, described briefly in Section 4[3]. In this Section, we present the classification results for only two real data sets, since for the other ones the conclusions of the results were very similar.

The first data set under consideration is the Canadian weather data set (see Figure 1), commonly used in the literature and available in the R package `fda` (Ramsay, Hooker, Graves, 2009; Ramsay et al., 2014). In this data set, the daily temperature and precipitation records of 35 Canadian weather stations averaged over 1960 to 1994 (365 days) are included. Thus, this data set contains 35 two-dimensional discrete functional observations observed on 350 design time points. These observations are assigned to one of two groups in a natural way. The first (resp. second) group consists of 5 Northern (resp. 30 Eastern and Western) weather stations located at higher latitudes (resp. at lower latitudes than these from the first group).

[2] It is worth noting that the outlying observations are present in all functional data sets considered. This was checked by the functional outlier detection method of Febrero-Bande, Galeano, González-Manteiga, (2007; 2008) implemented in the function `outliers.depth.trim()` available in the R package `fda.usc` (Febrero-Bande, Oviedo de la Fuente, 2012).

[3] The numerical experiments were performed in the R programming language (R Core Team, 2017). In this program, the implementations of the estimators in the functions `glm`, `glmRob` and `glmrob` from the packages `stats`, `robust` and `robustbase`, respectively, were used (Wang et al., 2014; Maechler et al., 2016).

As the second data set, we consider the ECG data set originated from Olszewski (2001) and investigated by Górecki, Krzyśko, Wołyński (2015) (see Figure 2). In this data set, two electrodes are used to collect data during one heartbeat. Each of 200 heartbeats is described by a two-dimensional discrete functional observation, and it is assigned to normal or abnormal group. Abnormal heartbeats are representative of a cardiac pathology, which is known as supraventricular premature beat. The normal (resp. abnormal) group consists of 133 (resp. 67) observations, which were observed at different design time points. For this reason, both discrete functional variables of this data set were extended to the same length of the longest one by using the method of Rodriguez, Alonso, Maestro (2005). The final common number of design points is 152.

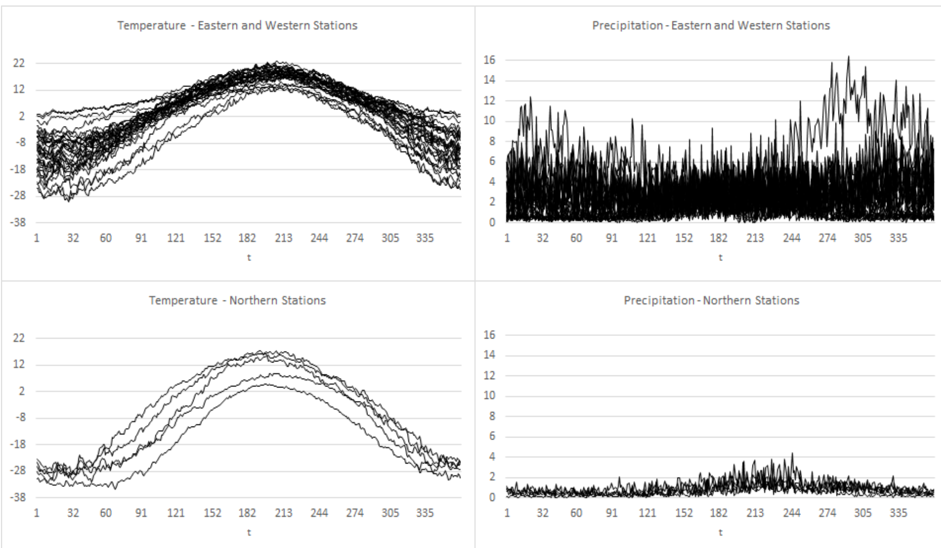

Figure 1. Canadian weather data

Source: the authors' research

Classifying the observations in both data sets is the binary classification problem for multi-dimensional functional data. The classifier (7) based on estimation methods described above is applied to this problem. The basis functions representation (4) of the observations was obtained by using the orthonormal Fourier basis and the least squares estimation method (see, for example, Krzyśko, Waszak 2013). For simplicity, equal truncation parameters for all variables were considered, i.e., $B_1 = B_2 = B$. More precisely, we only present the results for $B = 3, 5, \ldots, 13$ and $B = 3, 5, \ldots, 81$ for the Canadian weather and ECG data, respectively, since for greater values of B the classification error of all methods was very high (There was probably too many variables in the model (5) to obtain sensible estimation). Odd values of B are dictated by implementation of the Fourier basis in the R package `fda` (Ramsay, Hooker, Graves, 2009; Ramsay et al., 2014), which we used. Unfortunately, due to the low number of observations, the WBY estimator could not be used for the Canadian weather data set. This is to illustrate the limitation of the new methods, i.e., more observations may be needed to conduct the robust methods than for standard one.

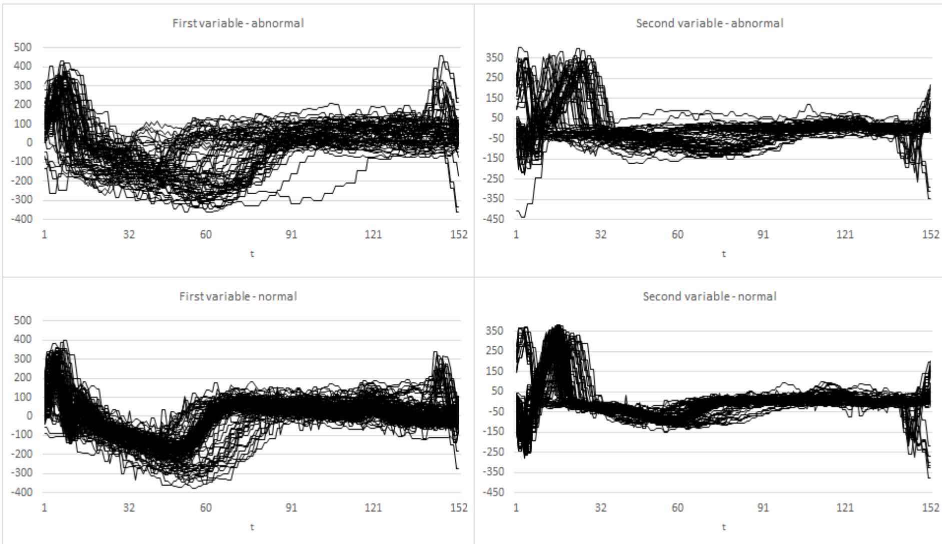

Figure 2. ECG data

Source: the authors' research

The 10-fold cross-validation method was used to calculate the classification error rates of the classifier (7) based on the MLE, MALLOWS, CUBIF, BY and WBY estimators. The results are depicted in Figures 3 and 4 for different values of truncation parameter B. They suggest that both functional data sets are quite difficult to recognize. Nevertheless, we can observe that the classifier (7) based on selected estimation techniques does not perform equally well. The robust estimators work at least as good as or even better than the MLE for most values of truncation parameter B. However, there is no method, which is superior in all situations. For different values of truncation parameter B, different estimation methods may classify best, e.g., for the ECG data and $B = 59$, the BY estimator works best, while for $B = 61$, the WBY one. For the Canadian weather data, the smallest 10-fold cross-validation error rate was achieved by the CUBIF estimator for $B = 5$, while for the ECG data, by the MALLOWS one for $B = 33$, 35. Therefore, for a given data set, the classifier (7) based on estimation methods under consideration as well as different bases and truncation parameters may be examined to select the method giving the smallest classification error.

6. Conclusions

We proposed the classification rule based on the functional logistic regression model with robust estimation methods of unknown parameters which leads to a novel solution of the classification problem of multivariate functional data. Numerical experiments for two real functional data sets indicate that the new methods usually work at least on par with the procedure of Górecki, Krzyśko, Wołyński (2015) and may be superior to it, especially in the presence of outlying observations. The proposed classifier is generic in nature, e.g., other choices of robust estimation methods in logistic regression model are also possible. The new classification rule can also be constructed by using non-orthogonal bases in contrast to that of Górecki, Krzyśko, Wołyński (2015). The appropriate choice of robust estimation method, basis functions, etc., should result in better performance of the proposed methods.

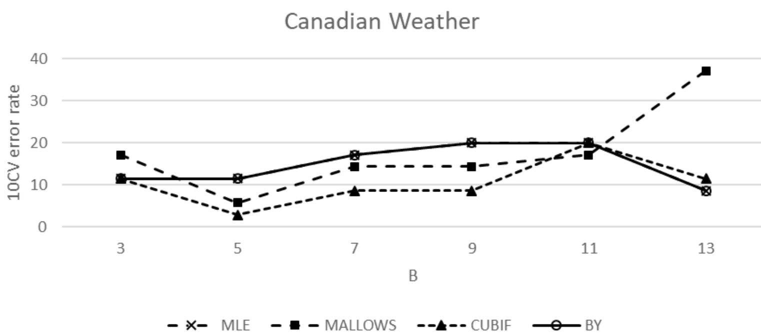

Figure 3. 10-fold cross-validation error rates (as percentages) for different values of truncation parameter *B* by using classifier (7) based on the MLE, MALLOWS, CUBIF and BY estimators for Canadian weather data

Source: the authors' research

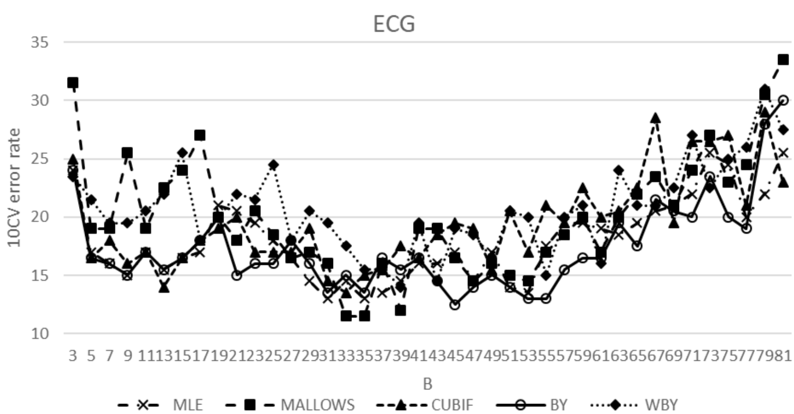

Figure 4. 10-fold cross-validation error rates (as percentages) for different values of truncation parameter *B* by using classifier (7) based on the MLE, MALLOWS, CUBIF, BY and WBY estimators for ECG data

Source: the authors' research

References

Ahmad S., Ramli N.M., Midi H. (2010), *Robust estimators in logistic regression: A comparative simulation study*, "Journal of Modern Applied Statistical Methods", vol. 9, pp. 502–511.

Bianco A.M., Yohai V.J. (1996), *Robust estimation in the logistic regression model*, [in:] H. Reider (ed.), *Robust statistics, Data analysis and computer intensive methods*, Springer Verlag, New York.

Chiou J.M., Müller H.G., Wang J.L. (2004), *Functional response models*, "Statistica Sinica", vol. 14, pp. 675–693.

Chiou J.M., Yang Y.F., Chen Y.T. (2016), *Multivariate functional linear regression and prediction*, "Journal of Multivariate Analysis", vol. 146, pp. 301–312.

Collazos J.A.A., Dias R., Zambom A.Z. (2016), *Consistent variable selection for functional regression models*, "Journal of Multivariate Analysis", vol. 146, pp. 63–71.

Croux C., Haesbroeck G. (2003), *Implementing the Bianco and Yohai estimator for logistic regression,* "Computational Statistics & Data Analysis", vol. 44, pp. 273–295.

Febrero-Bande M., Galeano P., González-Manteiga W. (2007), *A functional analysis of NO_x levels: location and scale estimation and outlier detection*, "Computational Statistics", vol. 22, pp. 411–427.

Febrero-Bande M., Galeano P., González-Manteiga W. (2008), *Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels*, "Environmetrics", vol. 19, pp. 331–345.

Febrero-Bande M., Oviedo de la Fuente M. (2012), *Statistical computing in functional data analysis: The R package fda.usc*, "Journal of Statistical Software", vol. 51, pp. 1–28.

Ferraty F., Vieu P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York.

Giacofci M., Lambert-Lacroix S., Marot G., Picard F. (2013), *Wavelet-based clustering for mixed-effects functional models in high dimension*, "Biometrics", vol. 69, pp. 31–40.

Górecki T., Krzyśko M., Wołyński W. (2015), *Classification problem based on regression models for multidimensional functional data*, "Statistics in Transition New Series", no. 16, pp. 97–110.

Górecki T., Łaźniewska E. (2013), *Funkcjonalna analiza składowych głównych PKB*, "Wiadomości Statystyczne", no. 4, pp. 23–34.

Górecki T., Smaga Ł. (2015), *A comparison of tests for the one-way ANOVA problem for functional data*, "Computational Statistics", vol. 30, pp. 987–1010.

Górecki T., Smaga Ł. (2017), *Multivariate analysis of variance for functional data*, "Journal of Applied Statistics", vol. 44, pp. 2172–2189.

Horváth L., Kokoszka P. (2012), *Inference for Functional Data with Applications*, Springer, New York.

Hubert M., Rousseeuw P.J., Segaert P. (2015), *Multivariate functional outlier detection*, "Statistical Methods & Applications", vol. 24, pp. 177–202.

James G.H., Hastie T.J. (2001), *Functional linear discriminant analysis for irregularly sampled curves*, "Journal of the Royal Statistical Society: Series B (Statistical Methodology)", vol. 63, pp. 533–550.

Jaworski S., Pietrzykowski R. (2014), *Spatial comparison of the level and rate of change of farm income in the years 2004–2012*, "Acta Universitatis Lodziensis, Folia Oeconomica", no. 307, pp. 29–44.

Kayano M., Konishi S. (2009), *Functional principal component analysis via regularized Gaussian basis expansions and its application to unbalanced data*, "Journal of Statistical Planning and Inference", vol. 139, pp. 2388–2398.

Krzyśko M., Waszak Ł. (2013), *Canonical correlation analysis for functional data*, "Biometrical Letters", no. 50, pp. 95–105.

Krzyśko M., Wołyński W. (2009), *New variants of pairwise classification*, "European Journal of Operational Research", vol. 199, pp. 512–519.

Krzyśko M., Wołyński W., Górecki T., Skorzybut M. (2008), *Learning Systems*, WNT, Warsaw.

Künsch H.R., Stefanski L.A., Carroll R.J. (1989), *Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models*, "Journal of American Statistical Association", vol. 84, pp. 460–466.

Maechler M., Rousseeuw P., Croux C., Todorov V., Ruckstuhl A., Salibian-Barrera A., Verbeke T., Koller M., Conceicao E.L.T., di Palma M.A. (2016), *robustbase: Basic Robust Statistics*, R package version 0.92–7, http://CRAN.R-project.org/package=robustbase [accessed: 5.04.2017].

Mallows C.L. (1975), *On some topics in robustness*, Bell Telephone Laboratories, Murray Hill.

Matsui H., Konishi K. (2011), *Variable selection for functional regression models via the L_1 regularization*, "Computational Statistics & Data Analysis", vol. 55, pp. 3304–3310.

Olszewski R.T. (2001), *Generalized feature extraction for structural pattern recognition in time-series data*. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, http://www.cs.cmu.edu/~bobski [accessed: 10.04.2017].

Ramsay J.O., Hooker G., Graves G. (2009), *Functional Data Analysis with R and MATLAB*, Springer, Berlin.

Ramsay J.O., Silverman B.W. (2002), *Applied Functional Data Analysis. Methods and Case Studies*, Springer, New York.

Ramsay J.O., Silverman B.W. (2005), *Functional Data Analysis*, 2nd Edition, Springer, New York.

Ramsay J.O., Wickham H., Graves S., Hooker G. (2014), *fda – Functional Data Analysis*, R package version 2.4.3, http://CRAN.R-project.org/package=fda [accessed: 28.01.2017].

R Core Team (2017), *R: A language and environment for statistical computing,* R Foundation for Statistical Computing, Vienna, https://www.R-project.org/ [accessed: 10.01.2017].

Rodriguez J.J., Alonso C.J., Maestro J.A. (2005), *Support vector machines of interval based features for time series classification*, "Knowledge-Based Systems", vol. 18, pp. 171–178.

Rousseeuw P.J. (1985), *Multivariate estimation with high breakdown point*, [in:] W. Grossmann, G. Pflug, I. Vincze, W. Wertz (eds.), *Mathematical Statistics and Applications*, vol. B, Reidel, Dordrecht.

Wang J., Zamar R., Marazzi A., Yohai V., Salibian-Barrera M., Maronna R., Zivot E., Rocke D., Martin D., Maechler M., Konis K. (2014), *robust: Robust Library*, R package version 0.4–16, https://CRAN.R-project.org/package=robust [accessed: 6.04.2017].

Zhang J.T. (2013), *Analysis of Variance for Functional Data*, Chapman & Hall, London.

Zastosowanie odpornej regresji logistycznej do klasyfikacji wielowymiarowych danych funkcjonalnych

Streszczenie: W niniejszym artykule rozważany jest problem dwuetykietowej klasyfikacji wielowymiarowych danych funkcjonalnych. Zaproponowane rozwiązanie tego problemu oparto na technikach regresyjnych i modelu regresji logistycznej dla danych funkcjonalnych. Model ten został przekształcony do szczególnego modelu regresji logistycznej za pomocą rozwinięcia (będących funkcjami) współczynników regresji i zmiennych objaśniających w bazie funkcyjnej. Na podstawie tego modelu skonstruowana została reguła klasyfikacyjna. W przypadku występowania obserwacji odstających rozważane są również metody odpornej estymacji nieznanych parametrów. Eksperymenty numeryczne sugerują, że proponowane metody mogą z powodzeniem być wykorzystane w praktycznych zagadnieniach.

Słowa kluczowe: analiza regresji dla danych funkcjonalnych, estymacja odporna, model regresji logistycznej, rozwinięcie funkcji w bazie funkcyjnej, wielowymiarowe dane funkcjonalne, zagadnienie klasyfikacji

JEL: C38, C13

Received: 2017-05-26; verified: 2017-11-23. Accepted: 2018-01-26