ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/236203892

The Mahalanobis Distance for Functional Data With Applications to Classification

Article *in* Technometrics · April 2013

DOI: 10.1080/00401706.2014.902774

CITATIONS
109

READS
570

3 authors, including:

Pedro Galeano
University Carlos III de Madrid
58 PUBLICATIONS **1,506** CITATIONS
SEE PROFILE

Rosa E. Lillo
University Carlos III de Madrid
168 PUBLICATIONS **1,597** CITATIONS
SEE PROFILE

All content following this page was uploaded by Rosa E. Lillo on 08 June 2016.

The user has requested enhancement of the downloaded file.

arXiv:1304.4786v1 [math.ST] 17 Apr 2013

The Mahalanobis distance for functional data with applications to classification

Esdras Joseph, Pedro Galeano and Rosa E. Lillo
Departamento de Estadística
Universidad Carlos III de Madrid

Abstract

This paper presents a general notion of Mahalanobis distance for functional data that extends the classical multivariate concept to situations where the observed data are points belonging to curves generated by a stochastic process. More precisely, a new semi-distance for functional observations that generalize the usual Mahalanobis distance for multivariate datasets is introduced. For that, the development uses a regularized square root inverse operator in Hilbert spaces. Some of the main characteristics of the functional Mahalanobis semi-distance are shown. Afterwards, new versions of several well known functional classification procedures are developed using the Mahalanobis distance for functional data as a measure of proximity between functional observations. The performance of several well known functional classification procedures are compared with those methods used in conjunction with the Mahalanobis distance for functional data, with positive results, through a Monte Carlo study and the analysis of two real data examples.

Keywords: Classification methods; Functional data analysis; Functional Mahalanobis semi-distance; Functional principal components.

1 Introduction

At the present time, there are a number of situations in different fields of applied sciences such as chemometrics, economics, image analysis, medicine, meteorology and speech recognition, among others, where it can be assumed that the observed data are points belonging to functions defined over a given set. Functional data analysis (FDA) deals with such kind of observations. In practice, the values of the functions are available only at a finite number of points and, as a general rule, functional samples may contain less functions than evaluation points. For this and other reasons, the majority of known multivariate tools can not be used for statistical analysis with this type of data since, by its nature, requires a different type of treatment. There are several methodologies for FDA being the most popular the one based on the use of basis functions such as Fourier and splines, see Ramsay and Silverman (2005). Alternatively, other procedures, such as the nonparametric approach proposed by Ferraty and Vieu (2006), do not require the knowledge of the explicit form of the functions. The ideas developed in this paper can be adapted to any of these situations. However, for easiness in exposition, the focus of this paper is on the basis functions approach.

Even if usual multivariate methods are not usually well suited for functional datasets, many multivariate techniques have inspired advances in FDA. The introduction of the notion of distance for functional data represents an example. Usually, it is assumed that the set of functions has been generated by a functional random variable defined on a Hilbert space endowed with a certain distance. However, in the recent literature on functional data, there is little reference to the role played by distances between functional data, with the book of Ferraty and Vieu (2006) an exception. These authors have proposed semi-metrics well adapted for sample functions, including semi-metrics based on functional principal components (FPC), partial least-squares (PLS) type semi-metrics and semi-metrics based on derivatives. However, common distances frequently used in multivariate data analysis such as the Mahalanobis distance proposed by Mahalanobis (1936) have not been extended to the functional framework. The first contribution of this paper is to fill this gap and presents the funcional Mahalanobis semi-distance that extends the multivariate Mahalanobis distance to the functional setting.

The use of distances in multivariate analysis is important in many different problems including classification, clustering, hypothesis testing and outlier detection, among others. In particular, several of the most well known methods for classification analysis are distance-based. Under a functional perspective, the aim of classification procedures is to decide

whether a function χ_0 generated from a functional random variable χ belongs to one of G classes using the information provided by G independent training samples $\chi_{g1}, \ldots, \chi_{gn_g}$, where $g = 1, \ldots, G$. Here χ_{gi}, for $i = 1, \ldots, n_g$, are independent replications of the functional random variable χ, measured on n_g randomly chosen individuals from class g. Using this information, a functional classification method provides a classification rule that can be used to classify χ_0. Nowadays, there is a wide variety of methods developed to solve this problem. For instance, several papers have proposed to classify functional observations by means of the functional principal component scores. For instance, Hall et al. (2001) proposed a method that consists in obtain the functional principal component scores of the training samples, then estimate nonparametrically the probability densities of the sets of functional principal component scores and finally estimate the posterior probability that χ_0 is of a given class using the Bayes classification rule. This approach was considered by Glendinning and Herbert (2003) for shape classification. Under a similar perspective, Leng and Müller (2006) proposed a method of classifying collections of temporal gene expression curves by means of functional logistic regression on the functional principal component scores of the training samples. Also, Song et al. (2008) compared several multivariate classification methods on the the basis expansion coefficients of the training samples for classifying time-course gene expression data. On the other hand, the popular nearest neighbor classification rule has been also extended to functional data. For instance, Biau et al. (2005) proposed to filter the training samples in the Fourier basis and to apply the kNN method to the first Fourier coefficients of the expansion, while Baíllo et al. (2011) derived several consistency results of the kNN procedure for a particular type of Gaussian processes. Additionally, the centroid method based on assign the function to the group with closer mean has been adapted to the functional framework by Delaigle and Hall (2012). Alternatively, several papers have extended the Fisher's discriminant analysis to the functional framework. The idea of these methods is to project the observations into a finite dimensional space where the classes are separated as much as possible. The transformed functions are called discriminant functions. Then, the new function χ_0 is also projected in this space and it is classified using the Bayes classification rule. In particular, James and Hastie (2001) used a natural cubic spline basis plus random error to model the observations from each individual. The spline is parameterized using a basis function multiplied by a coefficient vector, that is modeled using a Gaussian distribution. The observed functions can then be pooled to estimate the mean and covariance for each class by means of an Expectation-Maximization (EM) algorithm that are used to obtain the discriminant functions. Alternatively, Preda et al. (2007) used functional

PLS regression to obtain the discriminant functions while Shin (2008) considered an approach based on reproducing kernel Hilbert spaces. Finally, Ferraty and Vieu (2003) have proposed a method based on estimating nonparametrically the posterior probability that the new function χ_0 is of a given class, López-Pintado and Romo (2006), Cuevas et al. (2007) and Sguera et al. (2012) have proposed classifiers based on the notion of data depth that are well suited for datasets containing outliers, Rossi and Villa (2006) and Martin-Barragan et al. (2013) have investigated the use of support vector machines (SVMs) for functional data, Wang et al. (2007) have considered classification for functional data by Bayesian modeling with wavelet basis functions, Epifanio (2008) has developed classifiers based on shape descriptors, Araki et al. (2009) have considered functional logistic classification, and, finally, Alonso et al. (2012) have proposed a weighted distance approach. Note that, when a distance is required, these papers use the L^1, L^2 and L^∞ distances which are well defined in Hilbert spaces. The second contribution of this paper is to show that several simple classification procedures including the kNN procedure, the centroid method and functional Bayes classification rules can be used in conjunction with the functional Mahalanobis semi-distance as the criterion of proximity between functions to get very good classification rates without the need of much higher sophisticated classification methods. Several Monte Carlo experiments suggest that methods based on the functional Mahalanobis semi-distance leads to better classification rates than other alternatives.

The rest of this paper is organized as follows. Section 2 introduces the functional Mahalanobis semi-distance and shows some of its main characteristics. Section 3 reviews several classification methods for functional data and provides new approaches to these methods based on the functional Mahalanobis semi-distance. Section 4 analyzes the empirical properties of the procedures via several Monte Carlo experiments and illustrates the good behavior of the classification methods in conjunction with the functional Mahalanobis semi-distance through of the analysis of two real data examples. Finally, some conclusions are drawn in Section 5.

2 The funcional Mahalanobis semi-distance

2.1 Definitions and some characteristics

This section presents the functional Mahalanobis semi-distance that generalizes the Mahalanobis distance for multivariate random variables to the functional framework. Let $\mathbf{x}$ be a

multivariate continuous random variable defined in $\mathbb{R}^p$ with mean vector $\mathbf{m_x} = E[\mathbf{x}]$ and definite positive covariance matrix $\mathbf{C_x} = E\left[(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})'\right]$. The Mahalanobis distance between the random variable $\mathbf{x}$ and its mean vector $\mathbf{m_x}$ is the Euclidean norm of the random vector $\mathbf{C_x}^{-1/2}(\mathbf{x}-\mathbf{m_x})$ that can be written (see, Mahalanobis, 1936) as:

$$\begin{aligned} d_M(\mathbf{x},\mathbf{m_x}) &= \left\|\mathbf{C_x}^{-1/2}(\mathbf{x}-\mathbf{m_x})\right\|_E = \\ &= \left\langle \mathbf{C_x}^{-1/2}(\mathbf{x}-\mathbf{m_x}), \mathbf{C_x}^{-1/2}(\mathbf{x}-\mathbf{m_x})\right\rangle_E^{1/2} = \left[(\mathbf{x}-\mathbf{m_x})'\mathbf{C_x}^{-1}(\mathbf{x}-\mathbf{m_x})\right]^{1/2}, \end{aligned} \tag{1}$$

where $\|\cdot\|_E$ and $\langle\cdot,\cdot\rangle_E$ denote the Euclidean norm and the usual inner product in $\mathbb{R}^p$, respectively. The main characteristic of the multivariate Mahalanobis distance is that it takes into account the correlation structure of the multivariate random variable $\mathbf{x}$. Moreover, the multivariate Mahalanobis distance is scale invariant. For future developments, it is important to note that the Mahalanobis distance can be written in terms of the principal component scores of $\mathbf{x}$. For that, let $\mathbf{v}_1,\ldots,\mathbf{v}_p$ be the eigenvectors of the covariance matrix $\mathbf{C_x}$ associated with positive eigenvalues $a_1 \geq \cdots \geq a_p > 0$, and let $\mathbf{V}$ be the $p\times p$ matrix whose columns are the eigenvectors of the covariance matrix $\mathbf{C_x}$, i.e., $\mathbf{V} = [\mathbf{v}_1|\cdots|\mathbf{v}_p]$. Then, the vector of principal component scores given by $\mathbf{s} = \mathbf{V}'(\mathbf{x}-\mathbf{m_x})$, is a multivariate random variable with zero mean vector and diagonal covariance matrix. As a consequence, $\mathbf{x}$ can be written in terms of the principal component scores in the following way:

$$\mathbf{x} = \mathbf{m_x} + \mathbf{V}\mathbf{s}. \tag{2}$$

On the other hand, the singular value decomposition of $\mathbf{C_x}$, i.e., $\mathbf{C_x} = \mathbf{V}\mathbf{A}\mathbf{V}'$, where $\mathbf{A}$ is a diagonal matrix with the ordered eigenvalues $a_1,\ldots,a_p$ in the main diagonal, allows to write the inverse of $\mathbf{C_x}$ in terms of $\mathbf{V}$ and $\mathbf{A}$ as follows:

$$\mathbf{C_x}^{-1} = \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'. \tag{3}$$

Now, (2) and (3) leads to the following expression of the Mahalanobis distance between the random variable $\mathbf{x}$ and its mean vector $\mathbf{m_x}$ in terms of the principal component scores:

$$d_M(\mathbf{x},\mathbf{m_x}) = \left(\mathbf{s}'\mathbf{V}'\mathbf{V}\mathbf{A}^{-1}\mathbf{V}'\mathbf{V}\mathbf{s}\right)^{1/2} = \left(\mathbf{s}'\mathbf{A}^{-1}\mathbf{s}\right)^{1/2} = (\mathbf{z}'\mathbf{z})^{1/2}, \tag{4}$$

where $\mathbf{z} = \mathbf{A}^{-1/2}\mathbf{s}$ is the random vector of standardized principal component scores. In other words, the Mahalanobis distance between $\mathbf{x}$ and $\mathbf{m_x}$ can be written as the Euclidean norm

of the standardized principal component scores.

As mentioned before, the main goal of this section is to generalize the multivariate Mahalanobis distance to the functional setting. However, the proposal does not lead to a functional distance but to a functional semi-distance. The reasons of this will be clear once the functional Mahalanobis semi-distance is presented. For that, let χ be a functional random variable defined in the infinite dimensional space $L^2(T)$, i.e., the space of squared integrable functions in the closed interval $T = [a, b]$ of the real line. It is assumed that the functional random variable χ has a functional mean $\mu_\chi(t) = E[\chi(t)]$ and a covariance operator Γ_χ given by:

$$\Gamma_\chi(\eta) = E[(\chi - \mu_\chi) \otimes (\chi - \mu_\chi)(\eta)], \tag{5}$$

such that, for any $\eta \in L^2(T)$,

$$(\chi - \mu_\chi) \otimes (\chi - \mu_\chi)(\eta) = \langle \chi - \mu_\chi, \eta \rangle \, (\chi - \mu_\chi), \tag{6}$$

where $\langle ., . \rangle$ denotes the usual inner product on $L^2(T)$, i.e.:

$$\langle \chi - \mu_\chi, \eta \rangle = \int_T (\chi(t) - \mu_\chi(t)) \, \eta(t) dt.$$

The covariance operator Γ_χ in (5) is a well-defined compact operator so long as $E\left[\|\chi\|_2^4\right] < \infty$ (see Hall and Hosseini-Nasab, 2006), where $\|.\|_2$ denotes the usual norm in $L^2(T)$. Under this assumption, there exists a sequence of non-negative eigenvalues of Γ_χ, denoted by $\lambda_1 \geq \lambda_2 \geq \cdots$, where $\sum_{k=1}^{\infty} \lambda_k < \infty$, and a set of orthonormal eigenfunctions of Γ_χ, denoted by $\psi_1, \psi_2, \ldots$ such that $\Gamma_\chi(\psi_k) = \lambda_k \psi_k$, for $k = 1, 2, \ldots$ The eigenfunctions $\psi_1, \psi_2, \ldots$ form an orthonormal basis in $L^2(T)$ and allows to write the Karhunen-Loève expansion of the functional random variable χ (see Hall and Housseini-Nassab (2006)), in terms of the elements of the basis as follows:

$$\chi = \mu_\chi + \sum_{k=1}^{\infty} \theta_k \psi_k, \tag{7}$$

where $\theta_k = \langle \chi - \mu_\chi, \psi_k \rangle$, for $k = 1, 2, \ldots$ are the functional principal component scores of χ. It is well known that the functional principal component scores θ_k, for $k = 1, 2, \ldots$ are uncorrelated random variables with zero mean and variance λ_k since $\psi_1, \psi_2, \ldots$ are orthonormal.

In order to obtain a similar expression to (1) in the functional setting, it is necessary to define the inverse of the covariance operator, Γ_χ^{-1}. It exists under certain circumstances.

However, even in this case, Γ_χ^{-1} is unbounded and not continuous. Mas (2007) has proposed a regularized inverse operator which is a linear operator "close" to Γ_χ^{-1} and having good properties. For that, if Γ_χ^{-1} exists, this is given by:

$$\Gamma_\chi^{-1}(\zeta) = \sum_{k=1}^{\infty} \frac{1}{\lambda_k}(\psi_k \otimes \psi_k)(\zeta),$$

where ζ is a function in the range of Γ_χ. Then, the regularized inverse operator, denoted by Γ_K^{-1}, is defined as:

$$\Gamma_K^{-1}(\zeta) = \sum_{k=1}^{K} \frac{1}{\lambda_k}(\psi_k \otimes \psi_k)(\zeta),$$

where K is a given threshold. Similarly, it is also possible to give a regularized square root inverse operator given by:

$$\Gamma_K^{-1/2}(\zeta) = \sum_{k=1}^{K} \frac{1}{\lambda_k^{1/2}}(\psi_k \otimes \psi_k)(\zeta), \tag{8}$$

that allows to define the functional Mahalanobis semi-distance between χ and μ_χ inspired on (1) as follows:

Definition 2.1 *Let χ be a functional random variable defined in $L^2(T)$ with mean function μ_χ and compact covariance operator Γ_χ. The Mahalanobis semi-distance between χ and μ_χ, denoted by $d_{FM}^K(\chi, \mu_\chi)$, is defined as:*

$$d_{FM}^K(\chi, \mu_\chi) = \left\langle \Gamma_K^{-1/2}(\chi - \mu_\chi), \Gamma_K^{-1/2}(\chi - \mu_\chi) \right\rangle^{1/2}.$$

As noted before, the multivariate Mahalanobis distance may be expressed in terms of the principal component scores of the multivariate random variable $\mathbf{x}$. Similarly, it is possible to express the functional Mahalanobis semi-distance in terms of the functional principal component scores of the functional random variable χ as stated in the next proposition, that is proved in the appendix:

Proposition 2.1 *The functional Mahalanobis semi-distance between χ and μ_χ can be written as follows:*

$$d_{FM}^K(\chi, \mu_\chi) = \left(\sum_{k=1}^{K} \omega_k^2 \right)^{1/2}, \tag{9}$$

where $\omega_k = \theta_k/\lambda_k^{1/2}$*, for* $k = 1, \ldots, K$*, are the standardized functional principal component scores.*

Therefore, as in the multivariate case, the functional Mahalanobis semi-distance between χ and μ_χ is the Euclidean norm of the standardized functional principal component scores. This property provides a simple way to compute the functional Mahalanobis semi-distance in practice. It is also interesting to extend the definition of functional Mahalanobis semi-distance to the general situation of distance between two independent and identically distributed functional random variables.

Definition 2.2 *Let* χ_1 *and* χ_2 *be two functional random variables defined in* $L^2(T)$ *independent and identically distributed with mean function* μ_χ *and compact covariance operator* Γ_χ*. The functional Mahalanobis semi-distance between the functions* χ_1 *and* χ_2*, denoted by* $d_{FM}^K(\chi_1, \chi_2)$*, is given by:*

$$d_{FM}^K(\chi_1, \chi_2) = \left\langle \Gamma_K^{-1/2}(\chi_1 - \chi_2), \Gamma_K^{-1/2}(\chi_1 - \chi_2) \right\rangle^{1/2}.$$

The previous definition leads to the following proposition proven in the appendix:

Proposition 2.2 *The functional Mahalanobis semi-distance between* χ_1 *and* χ_2 *can be written as follow:*

$$d_{FM}^K(\chi_1, \chi_2) = \left(\sum_{k=1}^{K} (\omega_{1k} - \omega_{2k})^2 \right)^{1/2}, \tag{10}$$

where $\omega_{1k} = \theta_{1k}/\lambda_k^{1/2}$ *and* $\omega_{2k} = \theta_{2k}/\lambda_k^{1/2}$*, for* $k = 1, 2, \ldots$ *are the standardized functional principal component scores of* χ_1 *and* χ_2*, respectively.*

Therefore, the functional Mahalanobis semi-distance between two independent and identically distributed functional random variables can be written as the Euclidean distance between the standardized functional principal component scores of both functional random variables. The next result shows that d_{FM}^K is indeed a functional semi-distance.

Proposition 2.3 *Let* χ_1*,* χ_2 *and* χ_3 *be three independent and identically distributed functional random variables defined in* $L^2(T)$ *with mean function* μ_χ *and compact covariance operator* Γ_χ*. For any positive integer* K*,* d_{FM}^K *verifies the following three properties:*

1. $d_{FM}^K(\chi_1, \chi_2) \geq 0$.

2. $d^K_{FM}(\chi_1, \chi_2) = d^K_{FM}(\chi_2, \chi_1)$.

3. $d^K_{FM}(\chi_1, \chi_2) \leq d^K_{FM}(\chi_1, \chi_3) + d^K_{FM}(\chi_3, \chi_2)$.

Consequently, d^K_{FM} *is a functional semi-distance.*

It is well known that if the multivariate random variable $\mathbf{x}$ has a p-dimensional Gaussian distribution, then it is easy to see that $d^2_M(\mathbf{x}, \mathbf{m_x})$ has a χ^2_p distribution and, consequently, $E\left[d^2_M(\mathbf{x}, \mathbf{m_x})\right] = p$ and $V\left[d^2_M(\mathbf{x}, \mathbf{m_x})\right] = 2p$. To end this section, the following theorem shows a similar result for the functional Mahalanobis semi-distance.

Theorem 2.1 *If* χ *is a Gaussian process,* $d^K_{FM}(\chi, \mu_\chi)^2 \sim \chi^2_K$*, so that* $E\left[d^K_{FM}(\chi, \mu_\chi)^2\right] = K$ *and* $V\left[d^K_{FM}(\chi, \mu_\chi)^2\right] = 2K$.

2.2 Practical implementation

In practice, the functions are not observed continuously over all the points in the closed interval $T = [a, b]$, so that calculation of the functional Mahalanobis semi-distances as defined in (9) and (10) is not possible. Assume now that a dataset is observed with the following form:

$$\{\chi_i(t_{i,j}) : i = 1, \ldots, n \text{ and } j = 1, \ldots, J_i\}, \tag{11}$$

where n is the number of observed curves and J_i is the number of observations of the function χ_i at the points $t_{i,1}, \ldots, t_{i,J_i}$. Note that it is not assumed that the observation points are the same for all the functions not even their numbers. In this situation, the usual approach to obtain closed form expressions of the set of functions is to use basis functions. In general, a basis is a system of functions, denoted by ϕ_m, for $m = 1, 2, \ldots$, orthogonal or not, such that, for $i = 1, \ldots, n$:

$$\chi_i(t) \simeq \sum_{m=1}^{M} \beta_{im}\phi_m(t),$$

where β_{im}, for $m = 1, \ldots, M$, are the coefficients of the expansion. The number of basis functions, M, should be chosen on a case by case basis, although, M is usually chosen such that the functional approximations are close to the original counterparts with some smoothing that eliminates the most obvious noise. The choice of the basis is also important. There are several possibilities including polynomial, wavelets, Fourier and splines basis, among others. For periodic or nearly periodic datasets, Fourier basis is an adequate choice.

For nonperiodic datasets, B-splines are typically used. See Ramsay and Silverman (2005) for more information on basis functions. The simplest method to effectively estimate the coefficients of the expansion is carried out by minimizing:

$$\left(\sum_{j=1}^{J_i} \left[\chi_i \left(t_{i,j} \right) - \sum_{m=1}^{M} \beta_{im} \phi_m \left(t_{i,j} \right) \right]^2 \right)^{1/2}.$$

Now, with the smoothed functional sample, it is possible to estimate the functional mean μ_χ with the sample functional mean, $\widehat{\mu}_\chi$, given by:

$$\widehat{\mu}_\chi = \frac{1}{n} \sum_{i=1}^{n} \chi_i,$$

and the covariance operator Γ_χ with the sample covariance operator, $\widehat{\Gamma}_\chi(\eta)$, such that, for any $\eta \in L^2(T)$:

$$\widehat{\Gamma}_\chi(\eta) = \frac{1}{n} \sum_{i=1}^{n} \langle \chi_i - \widehat{\mu}_\chi, \eta \rangle (\chi_i - \widehat{\mu}_\chi).$$

Then, eigenfunctions and eigenvalues of the covariance operator Γ_χ can be approximated with those of $\widehat{\Gamma}_\chi$ leading to estimates $\widehat{\psi}_1, \widehat{\psi}_2, \ldots$ and $\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots$ respectively. Therefore, the functional principal component scores corresponding to curve χ_i, i.e., $\theta_{i,k} = \langle \chi_i - \mu_\chi, \psi_k \rangle$, are estimated with $\widehat{\theta}_{i,k} = \left\langle \chi_i - \widehat{\mu}_\chi, \widehat{\psi}_k \right\rangle$, for $k = 1, 2, \ldots$ that allows us to define the functional Mahalanobis semi-distance between χ_i and the functional sample mean $\widehat{\mu}_\chi$ as follows:

$$d_{FM}^K(\chi_i, \widehat{\mu}_\chi) = \left(\sum_{k=1}^{K} \widehat{\omega}_{ik}^2 \right)^{1/2},$$

where $\widehat{\omega}_{ik} = \widehat{\theta}_{i,k} / \widehat{\lambda}_k^{1/2}$, for $k = 1, \ldots, K$, are the sample standardized functional principal component scores. Similarly, the functional Mahalanobis semi-distance between two functions of the sample, χ_i and $\chi_{i'}$, can be written as follows:

$$d_{FM}^K(\chi_i, \chi_{i'}) = \left(\sum_{k=1}^{K} \left(\widehat{\omega}_{ik} - \widehat{\omega}_{i'k} \right)^2 \right)^{1/2},$$

where $\widehat{\omega}_{i'k} = \widehat{\theta}_{i',k} / \widehat{\lambda}_k^{1/2}$, for $k = 1, \ldots, K$.

3 Classification with the functional Mahalanobis semi-distance

Among all the possible applications of the functional Mahalanobis semi-distance introduced in the previous Section, this paper focuses in the supervised classification problem in the functional setting. Consider a sample of functional observations such that it is known in advance that each function comes from one of G predefined classes. Therefore, the whole sample can be split in G subsamples, denoted by $\chi_{g1}, \ldots, \chi_{gn_g}$, for $g = 1, \ldots, G$, respectively, where $n = n_1 + \cdots + n_G$ is the sample size of the whole dataset. Then, the idea is to use the information provided by the set of observations to construct classification rules that can be used to classify a new ungrouped functional observation χ_0. The aim of this section is to propose new procedures based on the combination of well known functional classification methods with the functional Mahalanobis semi-distance as a measure of proximity between functional objects. In particular, four procedures are presented.

3.1 The k-nearest neighbor (kNN) procedure

The k-nearest neighbor (kNN) procedure is one of the most popular methods used to perform supervised classification in multivariate settings. The method is very simple and appears to have a very good performance in many situations. Its generalization to infinite-dimensional spaces has been studied by Biau et al. (2005), Cérou and Guyader (2006) and Baíllo et al. (2011), among others. The kNN method starts by computing the distances between the new function to classify, χ_0, and all the functions in the observed sample. Next, the method finds the k functional observations in the sample closest in distance to χ_0. Finally, the new observation χ_0 is classified using majority of votes among the k neighbors. Cérou and Guyader (2006) have shown that the kNN procedure is not universally consistent. However, these authors have obtained sufficient conditions for consistency of the kNN classifier when the functional random variable takes values in a separable metric space. Additionally, Baíllo et al. (2011) have shown that the optimal classification rule can be explicitly obtained for a class of Gaussian processes with triangular covariance operators. The previous papers have considered three functional distances for the kNN classifier: the L^1, L^2 and L^∞ distances. In particular, the L^1, L^2 and L^∞ distances between χ_0 and the functional observation χ_{gi}

for $g = 1, \ldots, G$ and $i = 1, \ldots, n_g$ are given by:

$$d_1\left(\chi_0, \chi_{gi}\right) = \int_T \left|\chi_0\left(t\right) - \chi_{gi}\left(t\right)\right| dt,$$
$$d_2\left(\chi_0, \chi_{gi}\right) = \left(\int_T \left(\chi_0\left(t\right) - \chi_{gi}\left(t\right)\right)^2 dt\right)^{1/2},$$

and,

$$d_\infty\left(\chi_0, \chi_{gi}\right) = \sup\left\{\left|\chi_0\left(t\right) - \chi_{gi}\left(t\right)\right| : t \in T\right\},$$

respectively. Note that in order to compute the L^1, L^2 and L^∞ distances it is necessary to first smooth the discretized values of the function χ_0 as seen in Section 2.2. Also, it is important to note that no information about the class membership is used to compute the previous distances.

On the other hand, the kNN classifier can be used in conjunction with the functional Mahalanobis semi-distance. Contrary to the previous distances, two different ways to compute the functional Mahalanobis semi-distance in classification problems are in order. In a first case, assume that the functional means under class g, denoted by μ_{χ_g}, are different but the covariance operator, denoted by Γ_χ, is the same for all the classes. Then, the functional means, μ_{χ_g}, are estimated using the functional sample mean of the functions in class g, i.e.:

$$\widehat{\mu}_{\chi_g} = \frac{1}{n_g}\sum_{i=1}^{n_g} \chi_{gi}, \tag{12}$$

while the common covariance operator, Γ_χ, is estimated with the within class covariance operator given by:

$$\widehat{\Gamma}_\chi\left(\eta\right) = \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left\langle \chi_{gi} - \widehat{\mu}_{\chi_g}, \eta\right\rangle\left(\chi_{gi} - \widehat{\mu}_{\chi_g}\right), \tag{13}$$

for $\eta \in L^2\left(T\right)$. Now, the functional Mahalanobis semi-distance between χ_0 and the functional observation χ_{gi} for $g = 1, \ldots, G$ and $i = 1, \ldots, n_g$ is given by:

$$d_{FM}^K\left(\chi_0, \chi_{gi}\right) = \left(\sum_{k=1}^{K}\left(\widehat{\omega}_{g0k} - \widehat{\omega}_{gik}\right)^2\right)^{1/2}, \tag{14}$$

where $\widehat{\omega}_{g0k} = \widehat{\theta}_{g0k}/\widehat{\lambda}_k^{1/2}$ and $\widehat{\omega}_{gik} = \widehat{\theta}_{gik}/\widehat{\lambda}_k^{1/2}$, respectively, are the standardized sample functional principal component scores given by $\widehat{\theta}_{g0k} = \left\langle \chi_0 - \widehat{\mu}_{\chi_g}, \widehat{\psi}_k\right\rangle$ and $\widehat{\theta}_{gik} = \left\langle \chi_{gi} - \widehat{\mu}_{\chi_g}, \widehat{\psi}_k\right\rangle$,

respectively. Here, $\widehat{\psi}_1, \ldots, \widehat{\psi}_K$ and $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_K$ are the eigenfunctions and eigenvalues of the sample within class covariance operator (13). Similarly, the functional principal components (FPC) semi-distance proposed by Ferraty and Vieu (2006) between χ_0 and the functional observation χ_{gi} for $g = 1, \ldots, G$ and $i = 1, \ldots, n_g$, can be written as follows:

$$d_{FPC}^{K'}(\chi_0, \chi_{gi}) = \left(\sum_{k=1}^{K'} \left(\widehat{\theta}_{g0k} - \widehat{\theta}_{gik} \right)^2 \right)^{1/2}, \tag{15}$$

where K' is a certain threshold. In a second case, assume that both, the functional means and the covariance operators, denoted by Γ_{χ_g}, are different for the classes $1, \ldots, G$. Then, the functional means, μ_{χ_g}, are estimated using (12), while the covariance operator of each class is estimated using the functional sample covariance operator of the functions in class g, i.e.:

$$\widehat{\Gamma}_{\chi_g}(\eta) = \frac{1}{n_g} \sum_{i=1}^{n_g} \left\langle \chi_{gi} - \widehat{\mu}_{\chi_g}, \eta \right\rangle \left(\chi_{gi} - \widehat{\mu}_{\chi_g} \right), \tag{16}$$

for $\eta \in L^2(T)$. Now, the functional Mahalanobis semi-distance between χ_0 and the functional observation χ_{gi} for $g = 1, \ldots, G$ and $i = 1, \ldots, n_g$, is like in (14) but here $\widehat{\omega}_{g0k} = \widehat{\theta}_{g0k}/\widehat{\lambda}_{gk}^{1/2}$ and $\widehat{\omega}_{gik} = \widehat{\theta}_{gik}/\widehat{\lambda}_{gk}^{1/2}$, respectively, where $\widehat{\theta}_{g0k} = \left\langle \chi_0 - \widehat{\mu}_{\chi_g}, \widehat{\psi}_{gk} \right\rangle$, $\widehat{\theta}_{gik} = \left\langle \chi_{gi} - \widehat{\mu}_{\chi_g}, \widehat{\psi}_{gk} \right\rangle$ and $\widehat{\psi}_{g1}, \ldots, \widehat{\psi}_{gK}$ and $\widehat{\lambda}_{g1}, \ldots, \widehat{\lambda}_{gK}$ are the eigenfunctions and eigenvalues of the sample covariance operator in (16), respectively. Also, the FPC semi-distance in this second case can be written as in (15) but considering the same sample functional scores obtained with the eigenfunctions from the covariance operator (16), as before.

3.2 The centroid procedure

The centroid procedure for functional datasets, proposed by Delaigle and Hall (2012), is probably the fastest and simplest classification method for functional observations. The centroid method consists in assigning a new function χ_0 to the class with closer mean. Note that any functional distance can be used to implement the procedure. In particular, Delaigle and Hall (2012) considered the case of $G = 2$ classes that have different mean and a common covariance operator and proposed to project the functions into a given direction and then compute the squared Euclidean distance between the observations. More precisely, Delaigle and Hall (2012) proposed to use the centroid classifier with the distance between χ_0 and the

sample functional mean $\widehat{\mu}_{\chi_g}$, for $g = 1, 2$, denoted by DH, and given by:

$$d_{DH}\left(\chi_0, \widehat{\mu}_{\chi_g}\right) = \left| \sum_{k=1}^{K''} \widehat{\omega}_{0gk} \widehat{\delta}_{12k} \right|, \tag{17}$$

where K'' is a certain threshold, $\widehat{\omega}_{0gk}$ is computed as in the previous subsection assuming a common covariance operator and

$$\widehat{\delta}_{12k} = \frac{\left\langle \widehat{\mu}_{\chi_2} - \widehat{\mu}_{\chi_1}, \widehat{\psi}_k \right\rangle}{\widehat{\lambda}_k^{1/2}},$$

for $k = 1, \ldots, K''$.

Of course, other distances can be applied in the general case of G classes. In particular, the L^1, L^2 and L^∞ distances and the two versions of the functional Mahalanobis and functional principal components semi-distances introduced before, between χ_0 and the sample functional mean $\widehat{\mu}_{\chi_g}$, for $g = 1, \ldots, G$, can be used. In particular, the semi-distances are computed similarly in the previous Section but replacing χ_{gi} with $\widehat{\mu}_{\chi_g}$.

3.3 The functional linear and quadratic Bayes classification rules

In multivariate statistics, the Bayes classification rule is derived as follows. Let $\boldsymbol{x}$ be a p-dimensional continuous random variable and let $f_1, \ldots, f_G$ be the corresponding density functions of $\boldsymbol{x}$ under the G classes. Let $\pi_1, \ldots, \pi_G$ be the prior probabilities assigned to the G classes, verifying $\pi_1 + \cdots + \pi_G = 1$. Using the Bayes Theorem, the posterior probability that a new observation $\boldsymbol{x_0}$ generated from $\boldsymbol{x}$ comes from class g is given by:

$$P(g|\boldsymbol{x_0}) = \frac{\pi_g f_g(\boldsymbol{x_0})}{\pi_1 f_1(\boldsymbol{x_0}) + \cdots + \pi_G f_G(\boldsymbol{x_0})}, \tag{18}$$

respectively, where $P(1|\boldsymbol{x_0}) + \cdots + P(G|\boldsymbol{x_0}) = 1$. The Bayes rule classifies $\boldsymbol{x_0}$ in the class with largest posterior probability. In other words, $\boldsymbol{x_0}$ is classified in class g if $\pi_g f_g(\boldsymbol{x_0})$ is maximum. In particular, if the f_g densities are assumed to be Gaussian with different means $\mathbf{m}_{\boldsymbol{x}_g}$ but identical covariance matrix $\mathbf{C}_{\boldsymbol{x}}$, this is equivalent to classify $\boldsymbol{x_0}$ in class g if:

$$d_M\left(\boldsymbol{x_0}, \mathbf{m}_{\boldsymbol{x}_g}\right)^2 - 2\log \pi_g$$

is minimum, where $d_M\left(\boldsymbol{x_0}, \mathbf{m}_{\boldsymbol{x}_g}\right)^2 = \left(\boldsymbol{x_0} - \mathbf{m}_{\boldsymbol{x}_g}\right)' \mathbf{C}_{\boldsymbol{x}}^{-1} \left(\boldsymbol{x_0} - \mathbf{m}_{\boldsymbol{x}_g}\right)$ is the squared Mahalanobis distance between $\boldsymbol{x_0}$ and $\mathbf{m}_{\boldsymbol{x}_g}$.

Under the functional framework, the idea is to consider a similar rule but replacing the multivariate Mahalanobis distance with the functional Mahalanobis semi-distance. Consequently, assuming different means and a common covariance operator, the new observation χ_0 is assigned to the class g if:

$$d_{FM}^K\left(\chi_0, \widehat{\mu}_{\chi_g}\right)^2 - 2\log \pi_g \tag{19}$$

is minimum. Note that the values of π_g are usually fixed as the proportion of observations in the sample in the classes. In particular, if $\pi_1 = \cdots = \pi_G$, the linear Bayes classification rule reduces to the centroid classifier with the functional Mahalanobis semi-distance assuming a common covariance operator.

On the other hand, if in the multivariate case the f_g densities are assumed to be Gaussian with different means $\mathbf{m}_{\boldsymbol{x}_g}$ and different covariance matrices $\mathbf{C}_{\boldsymbol{x}_g}$, the Bayes rule classifies $\boldsymbol{x_0}$ in class g if:

$$d_M\left(\boldsymbol{x_0}, \mathbf{m}_{\boldsymbol{x}_g}\right)^2 + \log\left|\mathbf{C}_{\boldsymbol{x}_g}\right| - 2\log \pi_g$$

is minimum, where in this case, $d_M\left(\boldsymbol{x_0}, \mathbf{m}_{\boldsymbol{x}_g}\right)^2 = \left(\boldsymbol{x_0} - \mathbf{m}_{\boldsymbol{x}_g}\right)' \mathbf{C}_{\boldsymbol{x}_g}^{-1} \left(\boldsymbol{x_0} - \mathbf{m}_{\boldsymbol{x}_g}\right)$, is the squared Mahalanobis distance between $\boldsymbol{x_0}$ and $\mathbf{m}_{\boldsymbol{x}_g}$. Under the functional framework, the new observation χ_0 is assigned to the class G_0 if:

$$d_{FM}^K(\chi, \widehat{\mu}_{\chi_g})^2 + \sum_{k=1}^{K} \log(\widehat{\lambda}_{gk}) - 2\log \pi_g, \tag{20}$$

is minimum, where $\widehat{\lambda}_{gk}$, for $k = 1, \ldots, K$ are the eigenvalues of the estimated covariance operators under class g, respectively, and K is the number of eigenfunctions used to compute the functional Mahalanobis semi-distances.

It is important to note that although the functional linear and quadratic classification Bayes rules in (19) and (20) have been derived using the functional Mahalanobis semi-distance, these methods essentially consists in applying the multivariate linear and quadratic Bayes rules to the functional principal components scores, that are multivariate random variables. Hall et al. (2001) proposed to use the Bayes classification rule in (18) after estimating nonparametrically the density function of the functional principal components scores. However, these authors pointed out that a computationally less expensive method is

to use the multivariate quadratic Bayes classification rule which is essentially the rule given in (20).

4 Empirical results

This section illustrates the performance of the functional classification procedures presented in Section 3 through several Monte Carlo simulations using four different scenarios and the analysis of two real datasets.

4.1 Monte Carlo Study

The Monte Carlo study considers four different scenarios. The first scenario consists in two Gaussian processes defined in the closed interval $I = [0, 1]$, with different means, $\mu_1(t) = 20t^{1.1}(1-t)$ and $\mu_2(t) = 20t(1-t)^{1.1}$, respectively, and a common covariance operator with eigenfunctions $\psi_k(t) = \sqrt{2}\sin((k-0.5)\pi t)$ and associated eigenvalues $\lambda_k = 1/((k-0.5)\pi)^2$, for $k = 1, 2, \ldots$ Then, 1000 datasets are generated composed of n_1 functions from the first process and n_2 functions from the second process such that $n = n_1 + n_2$ is the whole sample size. The generated functions are observed at J equidistant points of the closed interval $I = [0, 1]$, where J is either 50 or 100. A Gaussian noise of variance 0.01 is added to each generated point. Then, once a dataset is generated in this way, the sample is split in a training sample and a test sample. The training sample is composed of n_{10} functions of the first process and n_{20} functions of the second process, while the test sample is composed of n_{11} functions of the first process and n_{21} functions of the second process such that $n_{10}+n_{11} = n_1$ and $n_{20} + n_{21} = n_2$, respectively. In particular, two different configurations are considered. In the first one, $n = 200$ with $n_1 = n_2 = 100$ and $n_{10} = n_{20} = 75$, respectively. In the second one, $n = 300$ with $n_1 = n_2 = 150$, and $n_{10} = n_{20} = 120$, respectively.

The second scenario is similar to the first one but the eigenvalues of the covariance operator are given by $\lambda_{1k} = 1/((k-0.5)\pi)^2$ and $\lambda_{2k} = 2/((k-0.5)\pi)^2$, for $k = 1, 2, \ldots$, for the first and second processes, respectively. Finally, the third and fourth scenarios are similar to the first and second ones but replacing the Gaussian process with a standardized exponential process with rate 1 and with the same mean functions and covariance operators. The discrete trajectories are converted to functional observations using a B-splines basis of order 6 with 20 basis functions that are enough to fit well the data. Figure 1 shows four datasets, once smoothing has been performed, corresponding to the four situations

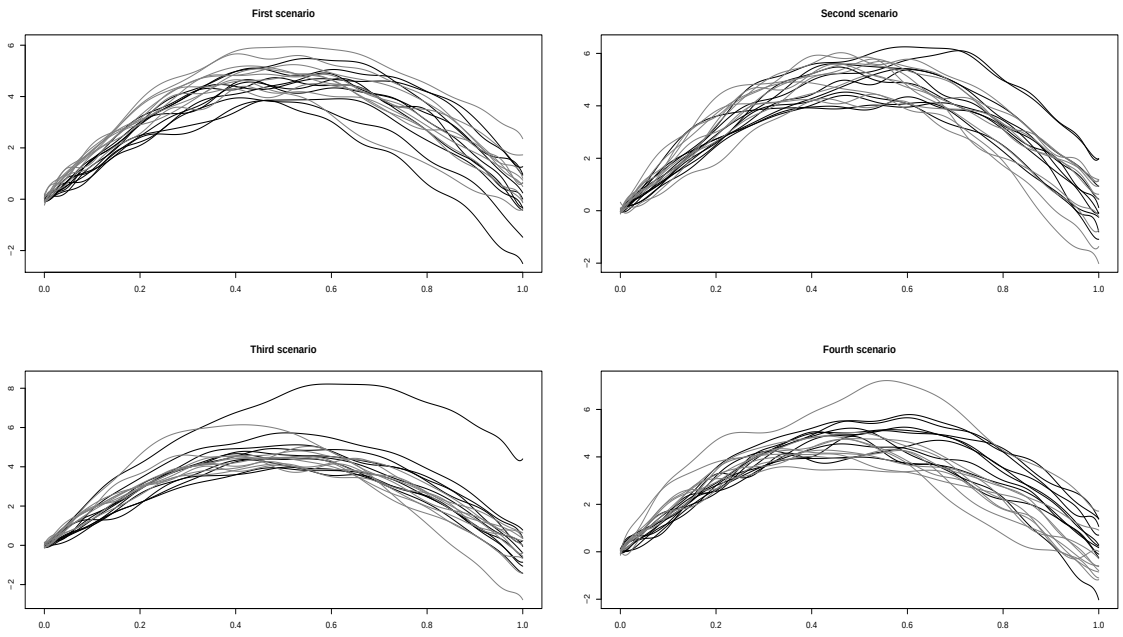

Figure 1: B-splines approximations of datasets corresponding to the four experiments considered. There are 10 functions per generated process.

considered. As it can be seen in the figure, the four scenarios appear to be complicated scenarios for classification purposes.

For each generated dataset, the functional observations in the test sample are classified using the following procedures: (1) the kNN procedure with seven different functional distances, the L_1, L_2 and L_∞ distances as proposed by Baíllo et al. (2011), the functional principal components (FPC) semi-distance assuming either a common or a different covariance operator, denoted by FPC_C and FPC_D, respectively, and the functional Mahalanobis (FM) semi-distance assuming either a common or a different covariance operator, denoted by FM_C and FM_D, respectively, as proposed in Section 3; (2) the centroid procedure with eight different functional distances, the first seven as in the kNN procedure and the distance proposed by Delaigle and Hall (2012) given in (17) and denoted by DH; (3) the linear and quadratic Bayes classification rules as proposed in Section 3, denoted by $FLBCR$ and $FQBCR$, respectively; and (4) the multivariate linear and quadratic Bayes classification rules applied on the coefficients of the B-splines basis representation, denoted by $LBCR$ $Coef.$ and $QBCR$ $Coef.$, respectively. This method can be seen as a simplification of the method proposed by James and Hastie (2001) much easier to implement than the original method. The threshold values needed to compute the FPC_C, FPC_D, FM_C, FM_D and

DH semi-distances and the $FLBCR$ and $FQBCR$ methods, and the maximum number of neighbors in the kNN procedures are determined using cross-validation with a maximum number of 15 eigenfunctions and 9 neighbors, respectively. Tables 1, 3, 5 and 7 show the proportion of correct classification of the test samples for the four scenarios. More precisely, each cell in the table displays the mean and the standard deviation (between parentheses) of the proportion of correct classifications over the 1000 Monte Carlo samples. On the other hand, Tables 2, 4, 6 and 8 show the means and standard deviations (between parentheses) of the optimal number of principal components needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods. In view of these tables, several comments are in order. First, in most of the cases, the kNN procedure with the FM_C semi-distance attains the largest proportion of correct classifications. Second, the proportions of correct classifications for the third and fourth scenarios are larger than the corresponding proportions for the first and second scenarios suggesting that Gaussianity is not necessarily an advantage for the functional Mahalanobis semi-distance. Third, in all the situations, classification methods in conjunction with the functional Mahalanobis semi-distance have a better performance than in conjunction with any other functional distance or semi-distance or any other alternative method as the one based on the basis functions coefficients. Fourth, there is not much difference in the results in terms of number of points in the grid and sample size. Fifth, at least in these scenarios, the use of the FPC_D and FM_D semi-distances is not of practical advantage. Indeed, even if the generated processes have different covariance operators, the methods appear to work better assuming a common covariance operator. Sixth, note that the multivariate quadratic Bayes classification rule for the coefficients of the Basis expansion has a bad performance in all the situations. This is probably due to the large amount of parameters that is necessary to estimate. Dimension reduction as done in James and Hastie (2001) may be a solution but at the cost of increasing the complexity of the procedure. In this sense note that very simple methods provides with very good performances without a high level of sophistication. Seventh, note also that, in most of the situations, standard deviations of good classification rates linked to method based on the functional Mahalanobis semi-distance are smaller than using any other alternative. Finally, note that there is no a clear pattern relative to the number of functional principal components used with the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances nor with the $FLBCR$ and $FQBCR$ methods. In summary, this limited simulation analysis appears to confirm that the functional Mahalanobis semi-distance may be a useful tool for classifying functional observations.

Table 1: Proportion of correct classification for the first scenario

n	J	**Method**	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
		kNN	.7657 (.0550)	.7655 (.0547)	.7682 (.0574)	.7866 (.0525)	.7871 (.0513)	.8314 (.0444)	.8209 (.0513)	–	–
		Centroid	.6710 (.0870)	.6823 (.0863)	.6764 (.0854)	.6868 (.0860)	.6907 (.0861)	**.8326** (.0490)	.8145 (.0480)	.8017 (.0570)	–
200	50	**FLBCR**	–	–	–	–	–	–	–	–	**.8326** (.0490)
		FQBCR	–	–	–	–	–	–	–	–	.8145 (.0480)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8201 (.0564)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7135 (.0708)
		kNN	.7700 (.0584)	.7721 (.0588)	.7744 (.0553)	.7924 (.0584)	.7918 (.0570)	**.8359** (.0463)	.8220 (.0570)	–	–
		Centroid	.6806 (.0920)	.6869 (.0832)	.6837 (.0806)	.6916 (.0817)	.6947 (.0824)	.8339 (.0542)	.8174 (.0539)	.8061 (.0625)	–
200	100	**FLBCR**	–	–	–	–	–	–	–	–	.8339 (.0542)
		FQBCR	–	–	–	–	–	–	–	–	.8174 (.0539)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8254 (.0552)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7255 (.0646)
		kNN	.7710 (.0523)	.7745 (.0524)	.7834 (.0490)	.7985 (.0496)	.7975 (.0510)	.8335 (.0452)	.8233 (.0510)	–	–
		Centroid	.6771 (.0794)	.6853 (.0752)	.6835 (.0729)	.6897 (.0761)	.6915 (.0762)	**.8350** (.0468)	.8239 (.0457)	.8049 (.0536)	–
300	50	**FLBCR**	–	–	–	–	–	–	–	–	**.8350** (.0468)
		FQBCR	–	–	–	–	–	–	–	–	.8239 (.0457)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8325 (.0488)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7660 (.0545)
		kNN	.7751 (.0529)	.7766 (.0521)	.7826 (.0520)	.7948 (.0523)	.7943 (.0492)	**.8378** (.0450)	.8225 (.0492)	–	–
		Centroid	.6935 (.0838)	.6906 (.0702)	.6925 (.0713)	.6936 (.0695)	.6966 (.0693)	.8348 (.0448)	.8231 (.0445)	.8063 (.0529)	–
300	100	**FLBCR**	–	–	–	–	–	–	–	–	.8348 (.0448)
		FQBCR	–	–	–	–	–	–	–	–	.8231 (.0445)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8290 (.0489)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7630 (.0545)

Table 2: Means and standard deviations (between parentheses) of the optimal number of principal components needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	6.36 (2.87)	6.53 (2.97)	7.48 (2.90)	7.06 (2.82)	–	–
Centroid	4.16 (2.09)	4.99 (2.52)	7.45 (2.99)	6.50 (2.87)	6.48 (3.06)	–
FLBCR	–	–	–	–	–	7.45 (2.99)
FQBCR	–	–	–	–	–	6.50 (2.87)
kNN	6.20 (2.63)	6.67 (2.68)	7.35 (2.85)	6.49 (2.83)	–	–
Centroid	4.05 (2.05)	4.66 (2.34)	7.36 (2.93)	6.32 (2.73)	6.86 (3.08)	–
FLBCR	–	–	–	–	–	7.36 (2.93)
FLBCR	–	–	–	–	–	6.32 (2.73)
kNN	6.57 (2.87)	6.69 (2.90)	7.40 (2.95)	7.21 (2.83)	–	–
Centroid	4.38 (2.17)	4.87 (2.49)	7.48 (3.11)	6.46 (2.87)	6.51 (3.00)	–
FLBCR	–	–	–	–	–	7.48 (3.11)
FQBCR	–	–	–	–	–	6.46 (2.87)
kNN	6.50 (2.73)	6.89 (2.87)	7.53 (2.78)	6.93 (2.87)	–	–
Centroid	4.39 (2.09)	4.61 (2.18)	7.83 (2.73)	6.67 (2.86)	6.94 (2.97)	–
FLBCR	–	–	–	–	–	7.83 (2.73)
FQBCR	–	–	–	–	–	6.67 (2.86)

Table 3: Proportion of correct classification for the second scenario

n	J	**Method**	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
		kNN	.7452 (.0556)	.7459 (.0555)	.7353 (.0543)	.7718 (.540)	.7718 (.0525)	**.8055** (.0474)	.7430 (.0525)	–	–
		Centroid	.6337 (.0783)	.6415 (.0774)	.6430 (.0734)	.6469 (.0774)	.6497 (.0783)	.7910 (.0549)	.7130 (.0554)	.7544 (.0754)	–
200	50	**FLBCR**	–	–	–	–	–	–	–	–	.7910 (.0549)
		FQBCR	–	–	–	–	–	–	–	–	.7130 (.0554)
		LBCR Coef.	–	–	–	–	–	–	–	–	.7753 (.0616)
		QBCR Coef.	–	–	–	–	–	–	–	–	.5817 (.0502)
		kNN	.7433 (.0550)	.7433 (.0516)	.7386 (.0481)	.7738 (.0494)	.7747 (.0469)	**.8058** (.0445)	.7407 (.0469)	–	–
		Centroid	.6350 (.0869)	.6439 (.0885)	.6513 (.0813)	.6531 (.0868)	.6566 (.0865)	.7928 (.0528)	.7207 (.0582)	.7578 (.0637)	–
200	100	**FLBCR**	–	–	–	–	–	–	–	–	.7928 (.0528)
		FQBCR	–	–	–	–	–	–	–	–	.7207 (.0582)
		LBCR Coef.	–	–	–	–	–	–	–	–	.7805 (.0581)
		QBCR Coef.	–	–	–	–	–	–	–	–	.5748 (.0514)
		kNN	.7543 (.0475)	.7550 (.0455)	.7500 (.0460)	.7833 (.0438)	.7825 (.0433)	**.8064** (.0397)	.7359 (.0433)	–	–
		Centroid	.6552 (.0714)	.6615 (.0705)	.6620 (.0752)	.6679 (.0695)	.6704 (.0708)	.7940 (.0494)	.7125 (.0548)	.7618 (.0550)	–
300	50	**FLBCR**	–	–	–	–	–	–	–	–	.7940 (.0494)
		FQBCR	–	–	–	–	–	–	–	–	.7125 (.0548)
		LBCR Coef.	–	–	–	–	–	–	–	–	.7897 (.0533)
		QBCR Coef.	–	–	–	–	–	–	–	–	.5604 (.0374)
		kNN	.7538 (.0490)	.7563 (.0510)	.7516 (.0490)	.7826 (.0488)	.7836 (.0492)	**.8098** (.0408)	.7385 (.0492)	–	–
		Centroid	.6499 (.0754)	.6601 (.0761)	.6651 (.0740)	.6650 (.0761)	.6681 (.0760)	.7967 (.0504)	.7097 (.0524)	.7650 (.0594)	–
300	100	**FLBCR**	–	–	–	–	–	–	–	–	.7967 (.0504)
		FQBCR	–	–	–	–	–	–	–	–	.7097 (.0524)
		LBCR Coef.	–	–	–	–	–	–	–	–	.7805 (.0490)
		QBCR Coef.	–	–	–	–	–	–	–	–	.5623 (.0401)

Table 4: Means and standard deviations (between parentheses) of the optimal number of principal components needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances for the second scenario and the $FLBCR$ and $FQBCR$ methods

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	6.46 (2.88)	6.11 (2.77)	6.11 (2.53)	4.94 (2.69)	–	–
Centroid	3.78 (2.00)	4.33 (2.49)	6.89 (2.98)	3.81 (1.75)	6.65 (3.08)	–
FLBCR	–	–	–	–	–	6.89 (2.98)
FQBCR	–	–	–	–	–	3.81 (1.75)
kNN	6.03 (2.75)	5.95 (2.71)	5.99 (2.44)	4.98 (2.53)	–	–
Centroid	3.77 (1.89)	4.38 (2.41)	6.96 (2.92)	3.62 (1.67)	6.52 (3.16)	–
FLBCR	–	–	–	–	–	6.96 (2.92)
FQBCR	–	–	–	–	–	3.62 (1.67)
kNN	6.40 (2.67)	6.33 (2.66)	5.95 (2.37)	4.76 (2.24)	–	–
Centroid	3.99 (1.97)	4.53 (2.47)	7.62 (2.93)	3.44 (1.20)	6.81 (3.11)	–
FLBCR	–	–	–	–	–	7.62 (2.93)
FQBCR	–	–	–	–	–	3.44 (1.20)
kNN	6.06 (2.91)	6.16 (2.74)	6.40 (2.51)	4.48 (2.26)	–	–
Centroid	4.07 (1.99)	4.82 (2.55)	7.41 (2.90)	3.32 (1.19)	7.10 (3.20)	–
FLBCR	–	–	–	–	–	7.41 (2.90)
FQBCR	–	–	–	–	–	3.32 (1.19)

Table 5: Proportion of correct classification for the third scenario

n	J	Method	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
		kNN	.8344 (.0498)	.8397 (.0504)	.8408 (.0439)	.8654 (.0445)	.8646 (.0437)	**.8999** (.0379)	.8842 (.0437)	–	–
		Centroid	.6738 (.0847)	.7052 (.0857)	.7131 (.0786)	.7093 (.0867)	.7128 (.0860)	.8392 (.0500)	.8165 (.0531)	.8099 (.0536)	–
200	50	**FLBCR**	–	–	–	–	–	–	–	–	.8392 (.0500)
		FQBCR	–	–	–	–	–	–	–	–	.8165 (.0531)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8264 (.0537)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7186 (.0649)
		kNN	.8409 (.0477)	.8453 (.0469)	.8444 (.0488)	.8669 (.0436)	.8670 (.0434)	**.9050** (.0365)	.8877 (.0434)	–	–
		Centroid	.6819 (.0927)	.7060 (.0944)	.7183 (.0917)	.7100 (0.954)	.7147 (.0949)	.8464 (.0512)	.8240 (.0513)	.8152 (.0574)	–
200	100	**FLBCR**	–	–	–	–	–	–	–	–	.8464 (.0512)
		FQBCR	–	–	–	–	–	–	–	–	.8240 (.0513)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8342 (.0574)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7259 (.0648)
		kNN	.8544 (.0408)	.8596 (.0409)	.8604 (.0407)	.8821 (.0368)	.8794 (.0364)	**.9086** (.0294)	.8984 (.0364)	–	–
		Centroid	.6957 (.0806)	.7227 (.0754)	.7228 (.0704)	.7280 (.0759)	.7310 (.0764)	.8484 (.0456)	.8317 (.0421)	.8223 (.0499)	–
300	50	**FLBCR**	–	–	–	–	–	–	–	–	.8484 (.0456)
		FQBCR	–	–	–	–	–	–	–	–	.8317 (.0421)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8415 (.0523)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7644 (.0576)
		kNN	.8570 (.0405)	.8640 (.0389)	.8621 (.0433)	.8864 (.0346)	.8861 (.0354)	**.9119** (.0338)	.9023 (.0354)	–	–
		Centroid	.7065 (.0868)	.7340 (.0834)	.7363 (.0721)	.7378 (.0838)	.7421 (.0834)	.8503 (.0461)	.8299 (.0455)	.8245 (.0526)	–
300	100	**FLBCR**	–	–	–	–	–	–	–	–	.8503 (.0461)
		FQBCR	–	–	–	–	–	–	–	–	.8299 (.0455)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8464 (.0495)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7623 (.0565)

Table 6: Means and standard deviations (between parentheses) of the optimal number of principal components needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances for the third scenario and the $FLBCR$ and $FQBCR$ methods

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	5.71 (2.84)	6.04 (2.74)	5.35 (2.59)	5.00 (2.45)	–	–
Centroid	4.25 (2.24)	4.73 (2.26)	7.20 (2.77)	6.09 (2.70)	6.42 (2.74)	–
FLBCR	–	–	–	–	–	7.20 (2.77)
FQBCR	–	–	–	–	–	6.09 (2.70)
kNN	5.62 (2.71)	5.98 (2.77)	5.14 (2.56)	4.92 (2.39)	–	–
Centroid	3.91 (1.98)	4.66 (2.43)	6.68 (2.76)	5.67 (2.60)	6.46 (3.00)	–
FLBCR	–	–	–	–	–	6.68 (2.76)
FQBCR	–	–	–	–	–	5.67 (2.60)
kNN	6.10 (2.84)	6.29 (2.89)	4.93 (2.37)	4.78 (2.26)	–	–
Centroid	4.22 (2.20)	4.84 (2.55)	7.14 (2.81)	6.04 (2.79)	7.03 (2.92)	–
FLBCR	–	–	–	–	–	7.14 (2.81)
FQBCR	–	–	–	–	–	6.04 (2.79)
kNN	5.91 (2.75)	6.52 (2.77)	4.58 (2.21)	4.52 (1.94)	–	–
Centroid	4.34 (1.99)	5.08 (2.40)	7.23 (2.97)	6.17 (2.85)	6.75 (2.91)	–
FLBCR	–	–	–	–	–	7.23 (2.97)
FQBCR	–	–	–	–	–	6.17 (2.85)

Table 7: Proportion of correct classification for the fourth scenario

n	J	**Method**	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
		kNN	.8645 (.0489)	.8647 (.0485)	.8517 (.0461)	.8851 (.0464)	.8830 (.0457)	**.9212** (.0349)	.8619 (.0457)	–	–
		Centroid	.7328 (.0849)	.7213 (.0892)	.6923 (.0903)	.7234 (.0897)	.7262 (.0887)	.8939 (.0424)	.8386 (.0478)	.8679 (.0447)	–
200	50	**FLBCR**	–	–	–	–	–	–	–	–	.8939 (.0424)
		FQBCR	–	–	–	–	–	–	–	–	.8386 (.0478)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8913 (.0450)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7173 (.0668)
		kNN	.8669 (.0465)	.8712 (.0447)	.8576 (.0490)	.8874 (.0414)	.8872 (.0425)	**.9223** (.0330)	.8599 (.0425)	–	–
		Centroid	.7345 (.0846)	.7291 (.0856)	.6969 (.0861)	.7316 (.0861)	.7337 (.0859)	.8930 (.0398)	.8335 (.0456)	.8681 (.0455)	–
200	100	**FLBCR**	–	–	–	–	–	–	–	–	.8930 (.0398)
		FQBCR	–	–	–	–	–	–	–	–	.8335 (.0456)
		LBCR Coef.	–	–	–	–	–	–	–	–	.8912 (.0412)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7077 (.0632)
		kNN	.8803 (.0407)	.8854 (.0419)	.8713 (.0430)	.9006 (.0383)	.8987 (.0387)	**.9265** (.0316)	.8670 (.0387)	–	–
		Centroid	.7303 (.0688)	.7250 (.0747)	.7040 (.0800)	.7268 (.0749)	.7293 (.0741)	.8947 (.0379)	.8340 (.0445)	.8647 (.0479)	–
300	50	**FLBCR**	–	–	–	–	–	–	–	–	.8947 (.0379)
		FQBCR	–	–	–	–	–	–	–	–	.8340 (.0445)
		LBCR Coef.	–	–	–	–	–	–	–	–	.9044 (.0389)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7270 (.0589)
		kNN	.8795 (.0420)	.8797 (.0442)	.8632 (.0427)	.8974 (.0386)	.8960 (.0392)	**.9279** (.0318)	.8710 (0.392)	–	–
		Centroid	.7346 (.0717)	.7289 (.0797)	.7015 (.0818)	.7314 (.0795)	.7330 (.0796)	.8936 (.0384)	.8324 (.0441)	.8649 (.0455)	–
300	100	**FLBCR**	–	–	–	–	–	–	–	–	.8936 (.0384)
		FQBCR	–	–	–	–	–	–	–	–	.8324 (.0441)
		LBCR Coef.	–	–	–	–	–	–	–	–	.9010 (.0402)
		QBCR Coef.	–	–	–	–	–	–	–	–	.7215 (.0613)

Table 8: Means and standard deviations (between parentheses) of the optimal number of principal components needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances for the fourth scenario and the $FLBCR$ and $FQBCR$ methods

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	6.41 (2.99)	6.51 (2.83)	6.49 (3.00)	3.72 (1.56)	–	–
Centroid	4.87 (2.51)	5.23 (2.72)	8.26 (2.65)	5.06 (2.49)	7.54 (2.77)	–
FLBCR	–	–	–	–	–	8.26 (2.65)
FQBCR	–	–	–	–	–	5.06 (2.49)
kNN	6.52 (2.83)	7.02 (2.79)	6.47 (2.94)	3.71 (1.73)	–	–
Centroid	5.07 (2.53)	5.51 (2.65)	8.32 (2.67)	4.62 (2.38)	7.77 (2.89)	–
FLBCR	–	–	–	–	–	8.32 (2.67)
FQBCR	–	–	–	–	–	4.62 (2.38)
kNN	6.96 (2.89)	7.11 (2.84)	6.16 (2.90)	3.50 (1.27)	–	–
Centroid	5.00 (2.39)	5.28 (2.31)	8.51 (2.64)	4.73 (2.14)	7.51 (2.90)	–
FLBCR	–	–	–	–	–	8.51 (2.64)
FQBCR	–	–	–	–	–	4.73 (2.14)
kNN	6.64 (2.72)	7.04 (2.85)	6.42 (2.94)	3.37 (1.09)	–	–
Centroid	4.90 (2.20)	5.55 (2.67)	8.42 (2.55)	4.96 (2.25)	7.30 (2.89)	–
FLBCR	–	–	–	–	–	8.42 (2.55)
FQBCR	–	–	–	–	–	4.96 (2.25)

4.2 Real data study: Tecator dataset

Next, the classification procedures are applied to the Tecator dataset previously considered by Ferraty and Vieu (2003), Rossi and Villa (2006), Li and Yu (2008), Alonso et al. (2012) and Martin-Barragan et al. (2013), among others. The dataset that consists of 215 near-infrared absorbance spectra of meat samples, recorded on a Tecator Infracted Food Analyzer is available at http://lib.stat.cmu.edu/datasets/tecator. The absorbance of a meat sample is a function given by $log_{10}(I_0/I)$ where I_0 and I are, respectively, the intensity of the light before and after passing through of the meat sample. Each observation consist of a 100-channel absorbance spectrum in the wavelength range 850-1050 nm, contents of moisture (water), fat and protein. Therefore, the recorded absorbance can be seen as a discretized version of the continuous process. The classification problem here is to separate meat samples with a high fat content (more than 20%) from samples with low fat content (less than 20%) based on the absorbance. Among the 215 samples, 77 have high fat content and 138 have low fat content. Previous analysis of this dataset have suggested that classification of the second order derivatives of the observed functions produces lower misclassification rates. Therefore, the analysis of the original data and their second order derivatives are carried out. In both cases, the discrete observations are converted to functional observations using a B-splines basis of order 6 with 20 and 40 basis functions, respectively, that are enough to fit well the data. Figure 2 shows the sample of these 100-channel absorbance spectrum and their second derivatives after smoothing.

In order to evaluate the performance of the functional classification methods given before, 1000 training samples are considered composed by 58 and 104 randomly chosen functions of meat with high fat content and low fat content, respectively. For each training sample, it is associated a test sample composed by the remaining 19 and 34 functions of meat with high fat content and low fat content, respectively. The classification results are shown in Tables 9 and 11 that show the mean and the standard deviation (between parentheses) of the proportion of correct classifications obtained via cross-validation for the two cases. As in the simulation study, the threshold values needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods, and the maximum number of neighbors in the kNN procedures are determined using cross-validation with a maximum number of 15 eigenfunctions and 9 neighbors, respectively. In both cases, the kNN procedure with the FM_C semi-distance is the winner. The highest proportions of correct classification for the Tecator dataset and the second order derivatives are 0.9835 and 0.9918, respectively, suggesting that it is not necessary to use the second order derivatives of the Tecator data to

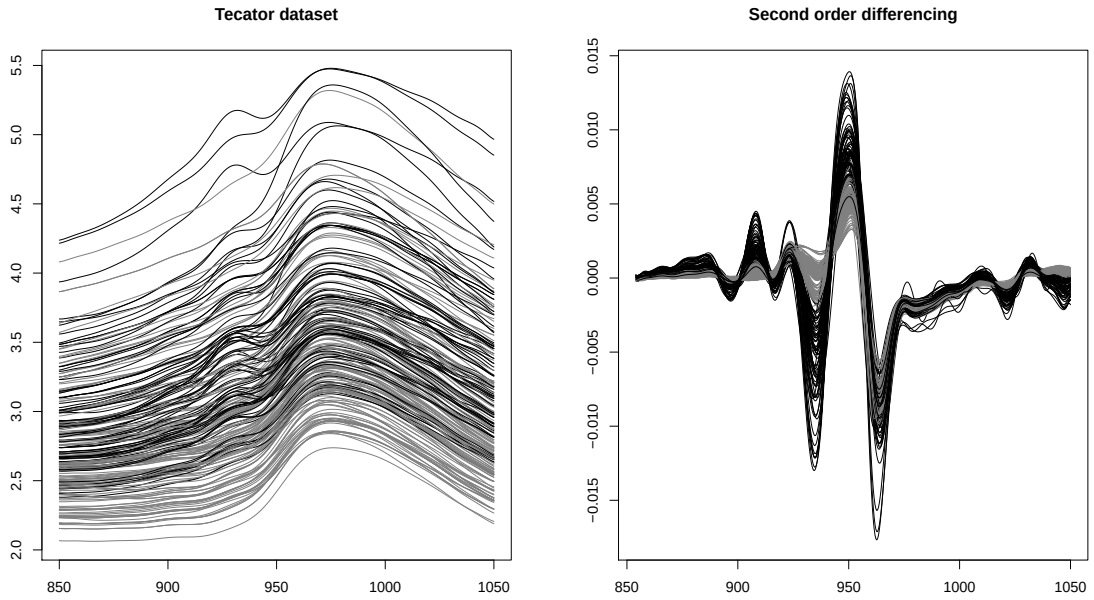

Figure 2: Right: Original observations of the Tecator dataset. Left: Second order derivatives of the Tecator dataset. High fat content in black and low fat content in gray

obtain almost perfect classification. Note that using a similar experiment, Rossi and Villa (2006) obtained good classification rates of 0.9672 and 0.9740 for the original and second order derivatives with SVMs, respectively, Li and Yu (2008) obtained good classification rates of 0.9602 and 0.9891 for the original and second order derivatives with a segmentation approach, respectively, Alonso et al. (2012) obtained good classification rates of 0.9798 and 0.9768, respectively, with two methods that takes into account the original, the first and the second order derivatives, and, finally, Martin-Barragan et al. (2013) obtained a good classification rate of 0.9891 with SVMs. Note that all of the previous approaches are more sophisticated than the ones taken here.

On the other hand, Tables 10 and 12 show the means and the standard deviation (between parentheses) of the number of principal components needed to calculate the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods for the original dataset and their second order derivatives. The mean numbers of functional principal components used with the functional Mahalanobis semi-distance are slightly larger than the corresponding to the functional principal components and Delaigle and Hall semi-distances if the original dataset is used but are sometimes smaller for their second order derivatives. Therefore, apparently there is not a general rule regarding the number of prin-

Table 9: Proportion of correct classification for the Tecator dataset

	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	.7904 (.0368)	.8108 (.0371)	.8602 (.0342)	.8144 (.0364)	.8135 (.0363)	**.9835** (.0114)	.9714 (.0363)	–	–
Centroid	.6784 (.0343)	.6812 (.0347)	.6957 (.0346)	.6813 (.0347)	.6813 (.0348)	.9630 (.0173)	.9521 (.0218)	.9479 (.0322)	–
FLBCR	–	–	–	–	–	–	–	–	.9517 (.0196)
FQBCR	–	–	–	–	–	–	–	–	.9671 (.0172)
LBCR Coef.	–	–	–	–	–	–	–	–	.9244 (.0244)
QBCR Coef.	–	–	–	–	–	–	–	–	.8958 (.0325)

Table 10: Means and standard deviations of the number of principal components used by the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods for the Tecator dataset

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	4.19 (.073)	4.89 (1.54)	4.86 (1.01)	5.12 (1.18)	–	–
Centroid	1.48 (0.98)	1.52 (1.09)	4.82 (0.94)	5.20 (1.24)	5.05 (1.47)	–
FLBCR	–	–	–	–	–	5.01 (1.10)
FQBCR	–	–	–	–	–	5.15 (1.10)

cipal components used.

4.3 Real data study: Phoneme dataset

Finally, the classification procedures are applied to the Phoneme dataset described in Ferraty and Vieu (2006) and available at http://www.math.univ-toulouse.fr/staph/npfda/npfda-datasets.html. The dataset contains log-periodograms corresponding to recordings of speakers of 32 ms duration. Here, two populations are considered corresponding to the phonemes "aa" as the vowel in "dark" and "ao" as the first vowel in "water", such that each speech frame is represented by 400 samples at a 16-kHz sampling rate where only the first 150 frequencies from each subject are retained. Therefore, the data consists of 800 log-periodograms of length 150, with known class phoneme membership. The classification problem here is to separate the two phonemes. The discrete observations are converted to functional observations using a B-splines basis of order 6 with 40 basis functions, respectively, that are enough

Table 11: Proportion of correct classification for the second order differences of the Tecator dataset

	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	.9885 (.0091)	.9852 (.0099)	.9814 (.0109)	.9901 (.0080)	.9870 (.0094)	**.9918** (.0076)	.9664 (.0094)	–	–
Centroid	.9629 (.0200)	.9608 (.0210)	.9546 (.0217)	.9651 (.0190)	.9617 (.0206)	.9678 (.0180)	.9372 (.0253)	.9630 (.0201)	–
FLBCR	–	–	–	–	–	–	–	–	.9533 (.0195)
FQBCR	–	–	–	–	–	–	–	–	.9555 (.0190)
LBCR Coef.	–	–	–	–	–	–	–	–	.9218 (.0261)
QBCR Coef.	–	–	–	–	–	–	–	–	.7220 (.0581)

Table 12: Means and standard deviations of the number of principal components used by the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods for the second order derivatives of the Tecator dataset

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	2.22 (0.62)	3.78 (1.99)	2.67 (1.83)	2.05 (1.00)	–	–
Centroid	1.63 (0.63)	3.35 (1.73)	2.99 (2.96)	1.66 (1.43)	3.59 (3.47)	–
FLBCR	–	–	–	–	–	4.24 (3.85)
FQBCR	–	–	–	–	–	2.00 (1.29)

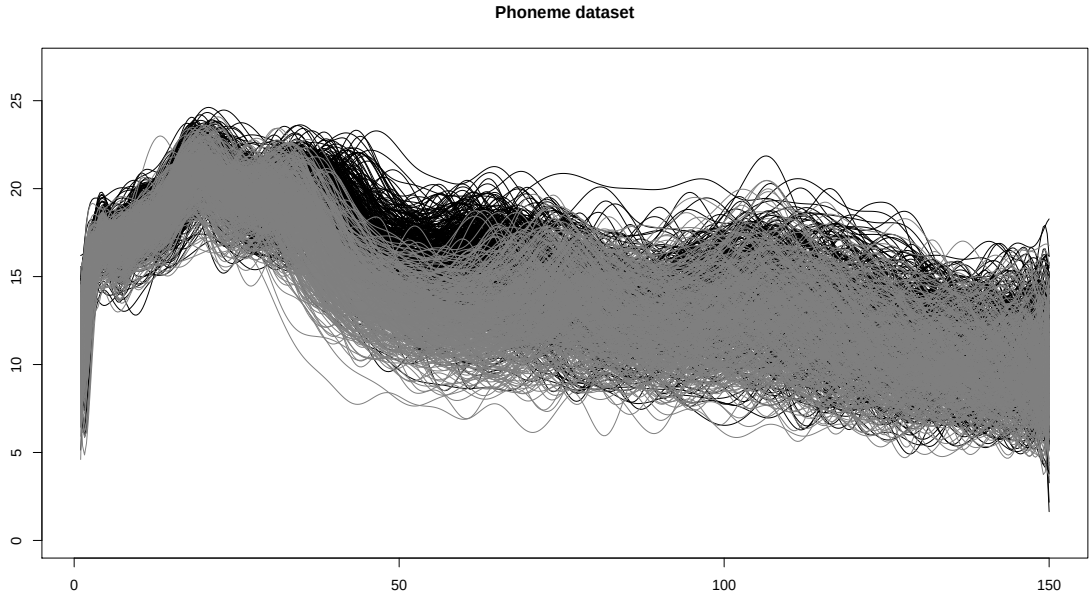

Figure 3: Phoneme dataset. Log-periodograms for "aa" in black and log-periodograms for "ao" in gray. Note that the log-periodograms in gray hide most of the log-periodograms in black

to fit well the data. Figure 3 shows the sample of log-periodograms. The figure confirms that it is difficult to distinguish the log-periodograms from one another.

As in the previous example, 1000 training samples are considered composed by 300 randomly chosen log-periodograms of both vowels. For each training sample, it is associated a test sample composed by the remaining 200 log-periodograms, 100 per vowel, respectively. The classification results are shown in Table 13 that shows the mean and the standard deviation (between parentheses) of the proportion of correct classifications obtained via cross-validation. As in the simulation study and the previous example, the threshold values needed to compute the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods, and the maximum number of neighbors in the kNN procedures are determined using cross-validation with a maximum number of 15 eigenfunctions and 9 neighbors, respectively. In this case, the centroid method with the FM_C semi-distance is the winner. Note that this method coincides in this case with the functional linear Bayes classification rule. The highest proportion of correct classification for the Phoneme dataset is 0.8238 which is slightly larger than other alternatives.

On the other hand, Table 14 shows the means and the standard deviation (between paren-

Table 13: Proportion of correct classification for the Phoneme dataset

	L^1	L^2	L^∞	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	.7918 (.0235)	.7847 (.0248)	.7838 (.0258)	.7996 (.0240)	.7799 (.0233)	.8124 (.0218)	.7961 (.0233)	–	–
Centroid	.7542 (.0319)	.7386 (.0307)	.7038 (.0283)	.7401 (.0307)	.7346 (.0303)	**.8238** (.0236)	.7994 (.0218)	.8001 (.0281)	–
FLBCR	–	–	–	–	–	–	–	–	**.8238** (.0236)
FQBCR	–	–	–	–	–	–	–	–	.7994 (.0218)
LBCR Coef.	–	–	–	–	–	–	–	–	.8050 (.0250)
QBCR Coef.	–	–	–	–	–	–	–	–	.7802 (.0261)

Table 14: Means and standard deviations of the number of principal components used by the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods with the Phoneme dataset

	FPC_C	FPC_D	FM_C	FM_D	DH	–
kNN	8.31 (2.87)	9.37 (3.26)	9.08 (2.82)	9.48 (3.40)	–	–
Centroid	6.83 (2.79)	8.68 (3.00)	8.94 (1.83)	8.04 (3.49)	8.08 (2.33)	–
FLBCR	–	–	–	–	–	8.94 (1.83)
FQBCR	–	–	–	–	–	8.04 (3.49)

theses) of the number of principal components needed to calculate the FPC_C, FPC_D, FM_C, FM_D and DH semi-distances and the $FLBCR$ and $FQBCR$ methods for the Phoneme dataset. The mean numbers of functional principal components used with the winner methods is around 9. However, other methods with worst performance have also mean values close to 9. Therefore, in this case, the differences between performances are apparently due to the methods themselves.

5 Conclusions

This paper has introduced a new semi-distance for functional data that generalize the multivariate Mahalanobis distance to the functional framework. For that, it is used the regularized square root inverse operator given in Mas (2007) that allows to write the functional Mahalanobis semi-distance between an observation and the sample mean function of the set of

functions in terms of the standardize functional principal component scores. Afterwards, new versions of several classification procedures have been proposed based on the functional Mahalanobis semi-distance. Some Monte Carlo experiments and the analysis of two real data examples illustrate the good behavior of the classification methods based on the functional Mahalanobis semi-distance. As mentioned previously, the range of applications of the functional Mahalanobis semi-distance is large and includes clustering, hypothesis testing and outlier detection, among others. This would be the objective of future work.

Acknowledgements

Financial support by MEC project ECO2012-38442 is gratefully acknowledged. The authors would like to thank Hugh Chipman for very helpful comments.

Appendix

Proof of Proposition 2.1

From (8), it is possible to write:

$$d_{FM}^K(\chi,\mu_\chi)=\left\langle \Gamma_K^{-\frac{1}{2}}(\chi-\mu_\chi),\Gamma_K^{-\frac{1}{2}}(\chi-\mu_\chi)\right\rangle^{1/2}=$$
$$=\left\langle \sum_{k=1}^K \frac{1}{\lambda_k^{1/2}}(\psi_k\otimes\psi_k)(\chi-\mu_\chi),\sum_{k=1}^K \frac{1}{\lambda_k^{1/2}}(\psi_k\otimes\psi_k)(\chi-\mu_\chi)\right\rangle^{1/2}.$$

Now, from (6) and (7), the previous expression leads to:

$$d_{FM}^K(\chi,\mu_\chi)=\left\langle \sum_{k=1}^K \frac{1}{\lambda_k^{1/2}}\left[\left\langle \psi_k,\sum_{j=1}^\infty \theta_j\psi_j\right\rangle\psi_k\right],\sum_{k=1}^K \frac{1}{\lambda_k^{1/2}}\left[\left\langle \psi_k,\sum_{j=1}^\infty \theta_j\psi_j\right\rangle\psi_k\right]\right\rangle^{1/2}.$$

As the inner product is linear for θ_k and the ψ_k are orthonormal eigenfunctions, it is possible to write:

$$d_{FM}^K(\chi,\mu_\chi)=\left\langle \sum_{k=1}^K \frac{\theta_k}{\lambda_k^{1/2}}\psi_k,\sum_{k=1}^K \frac{\theta_k}{\lambda_k^{1/2}}\psi_k\right\rangle^{1/2}=\left(\sum_{k=1}^K \frac{s_k^2}{\lambda_k}\right)^{1/2}=\left(\sum_{k=1}^K \omega_k^2\right)^{1/2}.$$

Proof of Proposition 2.2

By hypothesis, the two functions χ_1 and χ_2 have the same mean function, μ_χ, and the same covariance operator, Γ_χ. Therefore, from the Karhunen-Loève expansion:

$$\chi_1 = \mu_\chi + \sum_{k=1}^{\infty} \theta_{1k}\psi_k,$$

and,

$$\chi_2 = \mu_\chi + \sum_{k=1}^{\infty} \theta_{2k}\psi_k,$$

where $\theta_{1k} = \langle \chi_1 - \mu_\chi, \psi_k \rangle$ and $\theta_{2k} = \langle \chi_2 - \mu_\chi, \psi_k \rangle$, for $k = 1, \ldots$ are the functional principal component scores of χ_1 and χ_2, respectively. Consequently, the difference between the two functions χ_1 and χ_2 can be written as:

$$\chi_1 - \chi_2 = \sum_{k=1}^{\infty} (\theta_{1k} - \theta_{2k})\psi_k. \tag{21}$$

Using the expression (8) of the regularized square root inverse operator, the Mahalanobis semi-distance between χ_1 and χ_2 is given by:

$$d_{FM}^K(\chi_1, \chi_2) = \left\langle \Gamma_K^{-\frac{1}{2}}(\chi_1 - \chi_2), \Gamma_K^{-\frac{1}{2}}(\chi_1 - \chi_2) \right\rangle^{1/2} =$$

$$= \left\langle \sum_{k=1}^{K} \frac{1}{\lambda_k^{1/2}} (\psi_k \otimes \psi_k)(\chi_1 - \chi_2), \sum_{k=1}^{K} \frac{1}{\lambda_k^{1/2}} (\psi_k \otimes \psi_k)(\chi_1 - \chi_2) \right\rangle^{1/2}$$

Now, from (6) and (21), the above expression can be written as:

$$d_{FM}^K(\chi_1, \chi_2) = \left\langle \sum_{k=1}^{K} \frac{1}{\lambda_k^{1/2}} \langle \psi_k, \chi_1 - \chi_2 \rangle \psi_k, \sum_{k=1}^{K} \frac{1}{\lambda_k^{1/2}} \langle \psi_k, \chi_1 - \chi_2 \rangle \psi_k \right\rangle^{1/2} =$$

$$= \sum_{k=1}^{K} \frac{1}{\lambda_k} \left\langle \left\langle \psi_k, \sum_{j=1}^{\infty} (\theta_{1j} - \theta_{2j})\psi_j \right\rangle \psi_k, \left\langle \psi_k, \sum_{j=1}^{\infty} (\theta_{1j} - \theta_{2j})\psi_j \right\rangle \psi_k \right\rangle^{1/2} =$$

$$= \left(\sum_{k=1}^{K} \frac{1}{\lambda_k} \langle (\theta_{1k} - \theta_{2k})\psi_k, (\theta_{1k} - \theta_{2k})\psi_k \rangle \right)^{1/2} = \left(\sum_{k=1}^{K} (\omega_{1k} - \omega_{2k})^2 \right)^{1/2}$$

where $\omega_{1k} = \theta_{1k}/\lambda_k^{1/2}$ and $\omega_{2k} = \theta_{2k}/\lambda_k^{1/2}$, for $k = 1, 2, \ldots$ are the standardized functional principal component scores of χ_1 and χ_2, respectively.

Proof of Proposition 2.3

The proof of this proposition is trivial in view of Proposition 2.2 that asserts that $d_{FM}^K(\chi_1, \chi_2)$ is just the Euclidean distance between the first K standardized functional principal component scores of χ_1 and χ_2. Note that $d_{FM}^K(\chi_1, \chi_2)$ is not a functional distance because $d_{FM}^K(\chi_1, \chi_2) = 0$ if χ_1 and χ_2 have the same first K functional principal component scores, which does not imply $\chi_1 = \chi_2$.

Proof of Theorem 2.1

The functional Mahalanobis semi-distance between the Gaussian process χ and its mean function μ_χ is given in (9). Now, as χ is a Gaussian process, the standardized functional principal component scores, ω_k, for $k = 1, 2, \ldots$ are independent standard Gaussian random variables (see, Ash and Gardner, 1975) that shows the result.

References

Alonso, A. M., Casado, D. and Romo, J. (2012) Supervised classification for functional data: a weighted distance approach. Comput. Statist. Data Anal., 56, 2334-2346.

Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009) Functional logistic discrimination via regularized basis expansions. Commun. Statist. Theory. Math., 38, 2944-2957.

Ash, R. B. and Gardner, M. F. (1975) *Topics in stochastic processes*. Academic Press, New York.

Baíllo, A., Cuevas, A. and Cuesta-Albertos, J. A. (2011) Supervised classification for a family of Gaussian functional models. Scand. J. of Statist., 38, 480-498.

Biau, G., Bunea, F. and Wegkamp, M. H. (2005) Functional classification in Hilbert spaces. IEEE Trans. Inform. Theory., 51, 2163-2172.

Cérou, F. and Guyader, A. (2006) Nearest neighbor classification in infinite dimension. ESAIM: Probability and Statistics, 10, 340-355.

Cuevas, A., Febrero, M. and Fraiman, R. (2007) Robust estimation and classification for functional data via projection-based depth notions. Comput. Statist., 22, 481, 496.

Delaigle, A. and Hall, P. (2012) Achieving Near Perfect Classification for Functional Data. J. R. Statist. Soc. B, 74, 267-286.

Epifanio, I. (2008) Shape descriptors for classification of functional data. Technometrics, 50, 284-294.

Ferraty, F. and Vieu, P. (2003) Curves Discrimination: A Nonparametric Functional Approach. Comput. Statist. Data Anal., 51, 4878-4890.

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

Glendinning, R. H. and Herbert, R. A. (2003) Shape classification using smooth principal components. Pattern Recogn. Lett., 24, 2021-2030.

Hall, P., Poskitt, D. and Presnell, B. (2001) A functional data-analytic approach to signal discrimination. Technometrics, 43, 1-9.

Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis. J. R. Statist. Soc. B, 68, 109-126.

James, G. M. and Hastie, T. J. (2001) Functional linear discriminant analysis for irregularly sampled curves. J. R. Statist. Soc. B, 63, 533-550.

Leng, X. Y. and Müller, H.-G. (2006) Classification using functional data analysis for temporal gene expression data. Bioinformatics, 22, 68-76.

Li, B. and Yu, Q. (2008) Classification of Functional Data: A segmentation Approach. Comput. Statist. Data Anal., 52, 4790-4800.

López-Pintado, S. and Romo, J. (2006) *Depth-based classification for functional data*. In DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 72, 103-120. Providence: American Mathematical Society.

Mahalanobis, P. C. (1936) On the generalized distance in Statistics. Proc. Natl. Acad. Sci. India, 12, 49-55.

Martin-Barragan, B., Lillo, R. E. and Romo, J. (2013) Interpretable support vector machines for functional data. European J. of Oper. Res., In press.

Mas, A. (2007) Weak convergence in the functional autoregressive model. J. Mult. Anal., 98, 1231-1261.

Preda, C., Saporta, G. and Leveder, C. (2007) PLS classification of functional data. Comput. Statist., 22, 223-235.

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

Rossi, F. and Villa N. (2006) Support Vector Machine for Functional Data Classification.

Neurocomputing, 69, 730-742.

Shin, H. (2008) An extension of Fishers discriminant analysis for stochastic processes. J. Multiv. Anal., 99, 1191-1216.

Song, J. J., Deng, W., Lee, H.-J. and Kwon, D. (2008) Optimal classification for time-course gene expression data using functional data analysis. Comput Biol. Chem., 32, 426-432.

Wang, X. H., Ray, S. and Mallick, B. K. (2007) Bayesian curve classification using wavelets. J. Am. Statist. Ass., 102, 962-973.

View publication stats