Fuzzy Systems and Data Mining VI
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA200691

Curves Classification by Using a Local Likelihood Function and Its Practical Usefulness for Real Data

Mustapha Rachdi [a,1], Ali Laksaci [b], Ali Hamié [a], Jacques Demongeot [a] and Idir Ouassou [c]

[a] *Universityé Grenoble Alpes, France*
[b] *Department of Mathematics, College of Science and Statistical Research and Studies Support Unit, King Khalid University, Abha, KSA*
[c] *Université Cadi Ayyad (ENSAM) and Université Mohammed VI Polytechnique (Morocco)*

Abstract. We extend the classical approach in supervised classification based on the local likelihood estimation to the functional covariates case. The estimation procedure of the functional parameter (slope parameter) in the linear model when the covariate is of functional kind is investigated. We show, on simulated as well on real data, that classification error rates estimated using test samples, and the estimation procedure by local likelihood seem to lead to better estimators than the classical kernel estimation. In addition, this approach is no longer assuming that the linear predictors have a specific parametric form. However, this approach also has two drawbacks. Indeed, it was more expensive and slower than the kernel regression. Thus, as mentioned earlier, kernels other than the Gaussian kernel can lead to a divergence of the Newton-Raphson algorithm. In contrast, using a Gaussian kernel, 4 to 6 iterations are then sufficient to achieve convergence.

Keywords. Functional data analysis, Logit model, FPCA, Functional discrimination by local likelihood, NIR spectroscopic data, Mass spectroscopic data, Chromatography data, Waveform data

1. Introduction

A regular problem encountered in many scientific fields is the discrimination between curves. Generally, these curves describe the evolution of a quantity over time (monthly totals of precipitation, temperature evolution, patient walk curves with Parkinson's disease, ...) or when there are changes in absorbance depending on the wavelength (e.g., spectra provided by near infrared spectroscopy).
The collected data are in general considered as simple vectors of $\mathbb{R}^p$, but are curves obtained from observations at discretized random times of continuous functions of time. Traditionally, these observations are dependent on a time index which is a discretization

[1]Corresponding Author: University Grenoble Alpes, UFR SHS, BP. 47, BSHM, 38040 Grenoble Cedex 09, France E-mail: mustapha.rachdi@univ-grenoble-alpes.fr

grid. Thus, for each curve, the times of the grid in which the functions are observed, can be identical or different, uniformly distributed or not. Indeed, these data are in an infinite dimensional space, so they are called *Functional Data*.

In this work, we consider that each curve is associated with a qualitative variable admitting two terms, i.e., each curve is associated with a label. A classic example of label is "sick" or "healthy", "good" or "bad", 0 or 1. Therefore, the aim is to calculate the discrimination of curves. In other words, the goal is to explain the value of the label by the values of the curve in its entirety. Thus, another goal is the prediction: once a new curve is obtained, we predict the value that takes its label.

As part of the functional discrimination, most authors make some adaptations to extend the classical statistical models to functional case. We can mention, without being exhaustive, James and Hastie [1] applying the linear discriminant analysis of Fisher in case of functional variables. In 2002, James [2] offered the functional generalized linear model with a solution based on the EM algorithm (Expectation-Maximization) and Ferraty and Vieu [3] also offer non-parametric estimation methods of conditional probabilities based on kernel methods (see also [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] and [20] and references therein). Müller and Stadtmüller [21] propose the functional quasi-likelihood model. Moreover, Escabias *et al.* [22] and [23] offer to perform a logistic regression on principal components. Preda *et al.* [24] provide solutions to the wrong problem, the cover on infinite dimensional data, using partial least squares (PLS), and then, applying a linear analysis discriminant on the PLS components. Recently, Aguilera *et al.* [25] are interested in the same suite of PLS components used by Preda *et al.* [24], then they show that the PLS approach for functional data is equivalent to a multivariate PLS finished by using the coefficients of the approximations in basic functions as a predictor. Note that their method also applies in cases where the dependent variable is binary. Finally, Aguilera *et al.* [26] offer a functional logistic regression with three steps: first, functional data must be smoothed using penalized B-spline bases. Then, the principal components are extracted from the smoothed data. Finally, a logistic model is created using these components.

This paper is organized as follows. The first section presents the procedure of the functional principal components analysis generally used to reduce the dimension, and we recall some generalities about the generalized linear model. Then in Section 2, we briefly describe the approach of Ferraty and Vieu [3], with which we make comparisons to evaluate the performance of our approach. Then, in Section 3, we propose an extended functional version of multidimensional local likelihood. Section 4 focuses on implementation of the proposed method and that of Ferraty and Vieu [3] using real and simulated data.

2. Preliminaries on the eigenfunctions and the generalized linear model

Functional Principal Components Analysis (FPCA)

We consider, in what follows, that the functional variable X and $\mu_X(\cdot) = \mathbb{E}(X)(\cdot)$ exist and belong to $L^2[0,1]$. The covariance operator is given by

$$\Gamma_X(\eta) = \mathbb{E}[(X-\mu_X)\otimes(X-\mu_X)(\eta)],$$

such that, for all $\eta \in L^2[0,1]$

$$(X-\mu_X)\otimes(X-\mu_X)(\eta) = \langle X-\mu_X, \eta\rangle (X-\mu_X), \tag{1}$$

where $\langle\cdot,\cdot\rangle$ denotes the usual scalar product on $L^2[0,1]$.
The principal components analysis of X is based on the Karhunen-Loeve decomposition, breaking down the functional process on the basis of deterministic orthonormal functions. Indeed, there is a positive decreasing sequence $(\lambda_k)_{k\geq 1}$ of eigenvalues of Γ_X, where $\sum_{k=1}^{\infty}\lambda_k < \infty$, and an orthonormal family of functions $\psi_1, \psi_2, \ldots$, such that $\Gamma_X(\psi_k) = \lambda_k\psi_k,\ \ k = 1,2,\ldots$ This family of functions $\{\psi_k\}_{k\geq 1}$ forms an orthonormal base of functions in $L^2[0,1]$:

$$X = \mu_X + \sum_{k=1}^{\infty}\theta_k\psi_k, \tag{2}$$

where $\theta_k = \langle X-\mu_X, \psi_k\rangle,\ \ k = 1,2,\ldots$ are random coordinates (*functional principal component scores*), centered, of variance λ_k and not correlated. We can truncate this decomposition and keep only the $J > 1$ first terms.

3. Functional discrimination by local likelihood

As indicated above, our explanatory variables are assumed to be curves which are observed on all their trajectories. Thus, a natural generalization is obtained by replacing the finite sum by a definite integral on an infinite space [2],

$$g(\mathbb{E}[Y \mid X]) = \alpha + \int \beta(t)X(t)dt, \tag{3}$$

where $\alpha \in \mathbb{R}$ and where the $\beta(t)$ functions are assumed smooth and square integrable.
Thus, the goal is to model the relationship between a curve which corresponds to the functional variable X and a denoted membership to the class Y. We are in the context of an i.i.d. sample $(X_i, Y_i), i = 1,\ldots,n$, where $X \in L^2[0,1]$ and Y is of Bernoulli type.
The linear predictor is given by $\eta_i = \alpha + \int \beta(t)X_i(t)dt$, and accordingly, the generalized linear functional model is written in the form

$$Y_i = g^{-1}(\eta_i) + e_i, \quad i = 1,2,\ldots,n. \tag{4}$$

where $g(.)$ is the link function. The errors e_i are supposed to be independent and of zero mean. To simplify the notations, we set $g^{-1}(\eta_i) = \pi(X_i)$, where π assumed to be smooth enough, and is the inverse of g. So we define as part of a regression model $Y_i = \pi(X_i) + e_i$, $i = 1,\ldots n$. Then it is easy to write that

$$\pi(x) = \mathbb{E}(Y|X = x) = \mathbb{P}(Y = 1|X = x) \text{ with } 0 \leq \pi(x) \leq 1.$$

In the case of a dichotomous response variable, the likelihood associated with the sample size n is $\prod_{i=1}^{n}\pi(X_i)^{y_i}(1-\pi(X_i))^{y_i}$. Thus, the *overall log likelihood* is thus written:

$\mathscr{L}(\pi) = \sum_{i=1}^{n} \ell(y_i, \pi(X_i))$ where $\ell(y,\pi) = y\log\pi + (n-y)\log(1-\pi)$.
Unlike the generalized linear model, the approach of the local likelihood no longer assumes that η has a rigid parametric form. Assuming that η is a smooth function continuously differentiable, the idea is to approximate it locally by a polynomial of order 1 within a viewing window. It follows, via the Taylor expansion around x, that

$$\eta(X_i) \simeq \eta(x) + \langle \beta, X_i - x \rangle,$$

when X_i is located in a neighborhood of x, with $\eta(x)$ scalar noted α, $\beta = \beta(x) \in L^2[0.1]$ and $\langle \cdot, \cdot \rangle$ denoting the scalar product on $L^2[0,1]$. In the same way, the regression function π in X_i is approximated by the local logistic function:

$$\pi(X_i) \approx \frac{e^{\alpha + \langle \beta, X_i - x \rangle}}{1 + e^{\alpha + \langle \beta, X_i - x \rangle}}. \tag{5}$$

Moreover, the transformation logit is written, for $i = 1, \ldots, n$, as follows:

$$g(X_i) = \alpha + \langle \beta, X_i - x \rangle = \alpha + \int_{[0,1]} \big(\beta(t)(X_i(t) - x(t)\big) dt,$$

In order to estimate the model (6), we choose to adopt the local functional regression by Baíllo and Grané [28], whose response variable is scalar. Indeed, our minimization problem is:

$$\sum_{i=1}^{n} \big(g(X_i) - (\alpha + \langle \beta, X_i - x \rangle\big)^2 K\left(\frac{d(X_i,x)}{h}\right) \tag{6}$$

where K is a kernel whose role is to involve the variables X_i which belong to a ball centered at x and of radius h, also called the bandwidth h, and d is a semi-metric defining a measure of proximity between the curves, such that $d(X_i,x) \leq h$. The choice of a semi-metric d is discussed in Benhenni et al. [27], and the section 3.2 is devoted to the choice of the bandwidth h.

3.1. Estimation of the likelihood

In order to reduce the size of the parameter β, Baíllo and Grané [28] use an orthonormal basis $\{\phi_j\}_{1 \leq j \leq J}$ of dimension J over $L^2[0,1]$:

$$\beta = \sum_{j=1}^{J} \beta_j \phi_j \text{ and } X_i - x = \sum_{j=1}^{J} c_{ij} \phi_j,$$

with $\beta_j = \langle \beta, \phi_j \rangle$ and $c_{ij} = \langle X_i - x, \phi_j \rangle$.
So we choose our base ϕ by calculating the eigenfunctions of the empirical covariance operator (see Barrientos *et al.* [29]):

$$\frac{1}{|\mathscr{A}|} \sum_{i \in \mathscr{A}} (X_i - \overline{X})^t (X_i - \overline{X}), \text{ where } \mathscr{A} \text{is the learning sample.}$$

Therefore, for a fixed curve x, the experience planning matrix is written as follows:

$$\mathbf{X}_x = \begin{pmatrix} 1 & c_{11} & \cdots & c_{1J} \\ 1 & c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & & \vdots \\ 1 & c_{n1} & \cdots & c_{nJ} \end{pmatrix}.$$

Now, let's define a *local log-likelihood*. Indeed, the location is carried out via a nonnegative weighting function which depends on the distance between the curves:

$$\mathscr{L}_x(\alpha,\beta) = \sum_{i=1}^{n} K\Big(\frac{d(X_i,x)}{h}\Big)\Big(Y_i(\alpha + \langle \beta, X_i - x\rangle) - \log(1+e^{\alpha+\langle\beta,X_i-x\rangle})\Big), \qquad (7)$$

By passing to the matrix notation, we can write

$$\mathscr{L}_x(\gamma) = \mathbf{W}\ell(Y,\mathbf{X}_x\gamma), \qquad (8)$$

where $\mathbf{W}$ is a diagonal matrix $(n\times n)$ whose elements are the weights $K\left(\frac{d(X_i,x)}{h}\right)$ for $i=1,\ldots,n$ and $\gamma = (\alpha\ \beta_1 \ldots \beta_J)^\top$ is the vector of coefficients.
The solution of (8) is obtained by maximizing the log-likelihood. Generally, this solution is not analytical. Indeed, the optimal solution will be found by iterative methods, the most popular being the algorithms of Newton-Raphson and Fisher. The study of the maximum log-likelihood requires knowledge of derivatives. In the following, the first and second derivative of $\ell(y,\eta)$ with respect to η will be denoted respectively $\dot{\ell}(y,\eta)$ and $\ddot{\ell}(y,\eta)$.
So because we are part of a logistic regression, we can write the derivatives as follows:

$$\dot{\ell}(y,\eta) = y - h(\eta),\ \ddot{\ell}(y,\eta) = -h(\eta)(1-h(\eta)),\ \text{where } h(\eta) = e^{\eta}/(1+e^{\eta}).$$

The estimated parameter vector $\hat{\gamma}$ is therefore a solution for the local log-likelihood equation:

$$\mathbf{X}_x^\top \mathbf{W}\dot{\ell}(Y,\mathbf{X}_x\gamma) = 0 \qquad (9)$$

To find the local likelihood estimators at x fixed, we numerically solve the equation (9). Note that the system of equations (9) is convex, then it can be solved by the Newton-Raphson algorithm. Therefore, at iteration k, the estimator of Newton-Raphson is updated by the following equation

$$\gamma^{(k+1)} = \gamma^{(k)} + (\mathbf{X}_x^\top \mathbf{W}\mathbf{V}\mathbf{X}_x)^{-1}\mathbf{X}_x^\top \mathbf{W}\dot{\ell}(Y,\mathbf{X}_x\gamma^{(k)}), \qquad (10)$$

where $\mathbf{V}$ is a diagonal matrix whose diagonal elements are $-\big(\ddot{\ell}(Y_i,\mathbf{X}_x^i\gamma)\big)$.
Notice that the concavity of $\ddot{\ell}(Y_i,\mathbf{X}_x^i\gamma)$ implies that the matrix $\mathbf{X}_x^\top\mathbf{W}\mathbf{V}\mathbf{X}_x$ is positive definite, and that the matrix $\mathbf{X}_x^\top\mathbf{W}$ is full rank, which therefore causes the uniqueness of $\hat{\gamma}$.
Finally, once the parameters have been estimated, the estimator of $\pi(x)$ for x set is calculated using the equation (5):

$$\hat{\pi}(x) \simeq \frac{e^{\hat{\alpha}(x)}}{1+e^{\hat{\alpha}(x)}} \tag{11}$$

Therefore, the probability for the class $\{Y=0\}$ is given by:

$$\mathbb{P}(Y=0|X=x) = 1-\hat{\pi}(x).$$

Thus, a new curve x is assigned to the class $\{Y=1\}$ if $\mathbb{P}(Y=1|X=x) > 0.5$.
We note that, for among existing kernels, we used the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\}$. Indeed, other kernels provide smaller neighborhoods. Therefore, it is possible that all curves observed on a neighborhood (window) are from the same membership class. This then causes a divergence of the Newton-Raphson algorithm.

3.2. The bandwidth choice

The bandwidth parameter selection consists in applying a cross-validation procedure on a subset of smoothing parameters calculated from a k nearest neighbors (kNN) estimator. The idea is to consider neighborhoods whose size adapts automatically locally through a single parameter k. In other words, the bandwidth h is replaced by $h_k(x)$. More precisely, if $h_k(x)$ is the quantity associated for with exactly k curves $x_{i_1},\ldots,x_{i_k}$, such as $card\{i:\ d(x_i,x) < h_k(x)\} = k$, then $\pi_{g,k}^{(-i)}(x)$ is the leave-out one curve estimator of the local maximum likelihood dependency $h_k(x)$ built using a sample of the observed curves. Therefore, the optimal number k_{opt} of neighbors is defined as follows:

$$k_{opt} = \arg\min_k \sum_i \sum_{g=0}^{1} \left(\mathbb{1}_{[Y_i=g]} - \pi_{g,k}^{(-i)}(x)\right)^2 \tag{12}$$

Once the number of neighboring curves is optimized, then the selection process allows us to subsequently evaluate the estimator π in each curve x, by using the best local bandwidth $h(x) := h_{k_{opt}}(x)$.

4. Application on real and simulated data

In this section, we apply our method to three real data sets and a set of simulated data. In this way, we can compare our method with that of Ferraty and Vieu [3].

4.1. NIR spectroscopic data

Data were obtained from the analysis of samples of minced meat by near-infrared spectroscopy (NIR: *Near Infrared Spectroscopy*). Using this technology, the intensity of the absorption of near infrared ray and the wavelength (wavelengths between 850 and 1050 nanometers) have been measured. Thus, the chemical analysis was to evaluate the nutritional quality of meat.
There are 215 spectrometric curves corresponding to absorbance (equal to $-\log 10$(transmittance) measured by the apparatus) for 100 wavelengths evenly spaced between 850

and 1050 nm, of the corresponding fat. Moreover, these data can be regarded as random functions of the accomplishments: $X_i = \{X_i(\lambda), \lambda \in (850, 1050)\}$, for $i = 1, \ldots, 215$. Meat samples are divided into two classes: those that contain more than 20 % fat and those that contain less than 20 %. Figure 1 represents 20 curves for each class. The statistical discrimination allows avoiding chemical analysis, expensive and time consuming. These data, called Tecator data, and their detailed description are available on the site of StatLib[2].

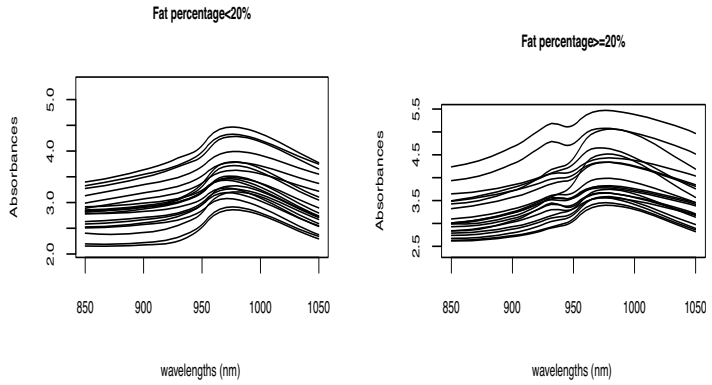

Figure 1. The two classes of data

In order to show the effectiveness of our method of estimation in the context of an anticipation, the sample is randomly divided between a learning subset noted $\mathscr{A}$ with $card(\mathscr{A}) = 160$, and a subset of test $\mathscr{T}$ with $card(\mathscr{T}) = 55$. Thus, the optimal number of neighbors k_{opt} will be calculated using $\mathscr{A}$, while $\mathscr{T}$ enables us to predict the classes of values $\hat{Y}_i$, $i \in \mathscr{T}$. Accordingly, the prediction quality is evaluated by the forecasting error (wrong-through rates) defined as follows:

$$\text{Error rate} = \frac{1}{|\mathscr{T}|} \sum_{i \in \mathscr{T}} \mathbb{1}_{[Y_i \neq \hat{Y}_i]}$$

We then apply the two methods, using a semi-metric based on derivatives of order 2 justified by the regular appearance of data. Thus, our method is evaluated on a grid of values of the parameter J representing the basis of size, to optimize within the meaning of the error rate. Therefore, we take $J = 5$ as the number of key components. After the experience of separation repeated 100 times data, the percentage of classification errors and the standard deviation associated to each method are given in Table 1. We will denote by KFV the classical kernel method in Ferraty and Vieu [37], and by LL the Local Likelihood method, showing the superiority of the second method.

4.2. Mass spectroscopic data

We consider a MALDI-TOF mass spectrometry dataset issued from a study on colorectal cancer (see Alexandrov *et al.* [30]). The sample set includes serum profiles of 64 subjects with colorectal cancer and 48 non-cancer control subjects. Each serum profile consists of 16331 recorded intensities corresponding to distinct m/z values.

Method	Mean error rate	Standard deviation
KFV	0.023	0.024
LL	0.018	0.016

Table 1. Tecator data: error rate on 100 test samples.

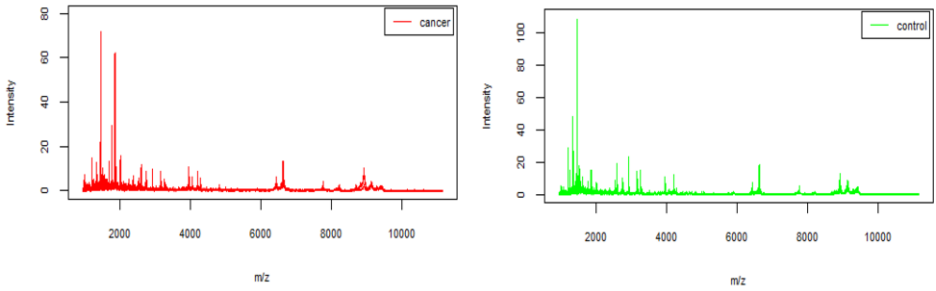

Figure 2. Example of spectra for each group coming from colorectal cancer dataset

Method	Mean error rate	Standard deviation
KFV	0.072	0.033
LL	0.060	0.034

Table 2. colorectal cancer: error rate on 100 test samples

The number of major components used to construct the semi-metric and reduce the size of the parameter β equals 4. Thus, Table 2 summarizes the results from 100 iterations, so that at each iteration the data is partitioned randomly into a training set of size 80 and a test set of size 32.

4.3. Chromatography Data derived from HPLC

The data comes from a study seeking to differentiate olive oil from several types of vegetable oils [31] and [32]. These data are composed of 115 oil samples analyzed by high performance liquid chromatography HPLC (*High-performance liquid chromatography*) coupled to an aerosol detector. Thus, the analysis provides 115 spectra of length 4001, also called chromatograms, 71 correspond to the olive oil and 44 are associated with other vegetable oils. The HPLC method provides profiles of triglycerides [3], which are a characteristic of different oils. Figure 3 represents 10 curves for each class (olive oil, vegetable oils). These data may be downloaded from the website: *http://www.models.life.ku.dk/oliveoil*. The data separation procedure is repeated 100 times, so that each training set of size 80, and each test set of size 35. Thus, a semi-metric

[2] *http ://lib.stat.cmu.edu/datasets/tecator*

[3] Triglycerides are molecules belonging to the class of lipids.

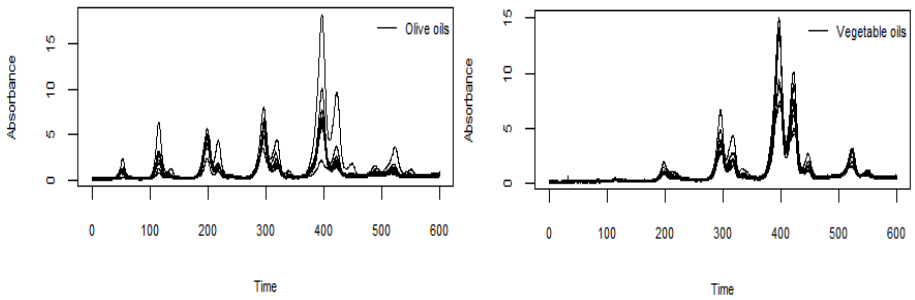

Figure 3. Oils analyzed by HPLC: 10 curves for each class

based on four main components is carried out. Moreover, the dimension J of the space base is also equal to 4. Finally, the results of the predictive performance of the two methods are summarized in the Table 3, showing anew the superiority of the second method.

Method	Mean error rate	Standard deviation
KFV	0.024	0.022
LL	0.008	0.013

Table 3. Data on Oils: error rate on 100 test samples

4.4. Simulated Data: "the waveform data"

As in Preda et al. [24], we plan to implement our method on simulated data, called Breiman *waveform*. This is a two class problem. Each class of curves is generated by linear combination of functions discretized into 101 points uniformly distributed in the interval $[1,21]$ and generated by the following equations:

$$\text{Class } \{Y=0\} : X(t) = Uh_1(t) + (1-U)h_2(t) + \varepsilon(t),$$
$$\text{Class } \{Y=1\} : X(t) = Uh_1(t) + (1-U)h_3(t) + \varepsilon(t),$$

where U is a uniform random variable on $[0,1]$, $\varepsilon(t)$ are independent and identically distributed normal standard variables, $h_1(t) = \max\{6-|t-11|,0\}, h_2(t) = h_1(t-4)$ and $h_3(t) = h_1(t+4)$. Figure 4 represents 20 curves for each class. The simulated curves are

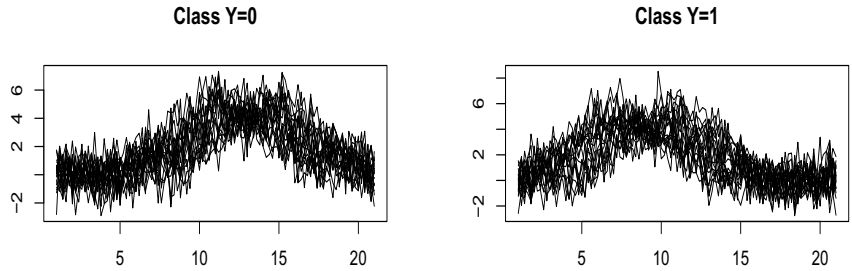

Figure 4. Simulated data: 20 curves for each class

randomly divided into two samples: a sample of 350 curves (175 per class), constituting

Method	Mean error rate	Standard deviation
KFV	0.040	0.022
LL	0.029	0.016

Table 4. *Waveform* data: error rate for 500 test samples

the learning base and a sample of 150 curves (75 per class) the test database. The optimum number of major components used to construct the semi-metric and the dimension of the projection base ϕ is 3. Thus, Table 4 gives the error rate resulting from the 50 first iterations.

Acknowledgment.

Prof. A. Laksaci extend his appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through General Research Project under grant number: G.R.P-90-41.

References

[1] G.M. James and T.J. Hastie (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, no. 3, Pages 533-550.

[2] G.M. James (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, no. 3, Pages 411-432.

[3] F. Ferraty and P. Vieu, "Curves discrimination: a nonparametric functional approach," *Computational Statistics & Data Analysis*, vol. 44, no. 1, pp. 161-173, 2003.

[4] A.M. Almanjahie, Z.C. Elmezouar, A. Laksaci, M. Rachdi (2019). FDA: strong consistency of the kNN local linear estimation of the functional conditional density and mode. Journal of Nonparametric Statistics, Vol. 31, no. 1, Pages 175-195.

[5] A.M. Almanjahie, Z.C. Elmezouar, A. Laksaci, M. Rachdi (2018). kNN local linear estimation of the conditional cumulative function: Dependent functional data case. C. R. Acad. Sci. Paris, Mathématiques, Sér. I, Vol. 356, no. 10, Pages 1036-1039.

[6] B. Altendji, J. Demongeot, A. Laksaci, M. Rachdi (2018). Functional Data Analysis: Estimation of the relative error in functional regression under random left-truncation. Journal of Nonparametric Statistics, Vol. 30, no. 2, Pages 472-490.

[7] F. Awadhi, Z. Kaid, A. Laksaci, I. Ouassou, M. Rachdi (2019). Functional data analysis: local linear estimation of the L1-conditional quantiles. Statistical Methods & Applications, Vol. 28, no. 2, Pages 217-240.

[8] K. Benhenni, S. Hedli-Griche and and M. Rachdi (2017). Regression models with correlated errors based on functional random design. TEST, Vol. 26, no. 1, Pages 1-21.

[9] O. Bouanane, A. Laksaci, M. Rachdi, S. Rahmani (2018). Asymptotic normality of some conditional nonparametric functional parameters in high-dimensional statistics. Behaviormetrika, Vol. 46, no. 1, Pages 199-233.

[10] J. Demongeot, A. Hamié, A. Laksaci, M. Rachdi (2015). Relative-error prediction in nonparametric functional statistics: theory and practice. Journal of Multivariate Analysis, Vol. 146, Pages 261-268.

[11] J. Demongeot, A. Laksaci, F. Madani, M. Rachdi (2011). *A fast functional locally modeled conditional density and mode for functional time-series*. Recent Advances in Functional Data Analysis and Related Topics, Contributions to Statistics, 2011, Pages 85-90, Physica-Verlag/Springer.

[12] J. Demongeot, A. Laksaci, F. Madani and M. Rachdi (2010). Estimation locale linéaire de la densité conditionnelle pour des données fonctionnelles. C. R. Acad. Sci. Paris, Sér. I, 348, No. 15-16, Pages 931-934.
[13] J. Demongeot, A. Laksaci, A. Naceri and M. Rachdi (2016). Estimation locale linéaire de la fonction de régression pour des variables hilbertiennes C. R. Mathématiques, Vol. 354, Issue 8, Pages 847-850.
[14] J. Demongeot, A. Laksaci, M. Rachdi and S. Rahmani (2014). On the local linear modelization of the conditional distribution for functional data Sankhyä A: The indian journal of statistics, Vol. 76, Issue 2, Pages 328-355.
[15] M. El Methni and M. Rachdi (2011). Local weighted average estimation of the regression operator for functional data. Communications in Statistics - Theory and Methods, Volume 40, Issue 17, Pages 3141-3153.
[16] S. Dabo-Niang, M. Rachdi and A.-F. Yao (2011). Kernel regression estimation for spatial functional random variables. Far East Journal of Theoretical Statistics, Volume 37, Issue 02, Pages 77-113.
[17] L. Kara-Zaitri, A. Laksaci, M. Rachdi and P. Vieu (2017). Data-driven kNN estimation for various problems involving functional data. Journal of Multivariate Analysis, vol. 153, Pages 176-188.
[18] F. Messaci, N. Nemouchi, I. Ouassou and M. Rachdi (2015). Local polynomial modelling of the conditional quantile for functional data Statistical Methods & Applications, vol. 29, no. 1, Pages 597-622.
[19] I. Ouassou and M. Rachdi (2012). Regression operator estimation by delta-sequences method for functional data and its applications. AStA Advances in Statistical Analysis, 2012, vol. 96, Pages 451-465.
[20] M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali and F. Madani (2014). Theoretical and practical aspects on the quadratique error in the local linear estimation of the conditional density for functional Computational Statistics & Data Analysis, Vol. 73, Issue 2, Pages 53-68.
[21] H.G. Müller and U. müller, "Generalized functional linear models," *Annals of Statistics*, 774-805, 2005.
[22] M. Escabias, A. M. Aguilera, and M. J. Valderrama, "Principal component estimation of functional logistic regression: discussion of two different approaches," *Journal of Nonparametric Statistics*, vol. 16, no. 3-4, pp. 365-384, 2004.
[23] M. Escabias, A. M. Aguilera, and M. J. Valderrama, "Modeling environmental data by functional principal component logistic regression," *Environmetrics*, vol. 16, no. 1, pp. 95-107, 2005.
[24] C. Preda, G. Saporta, and C. Lévéder, "PLS classification of functional data," *Computational Statistics*, vol. 22, no. 2, pp. 223-235, 2007.
[25] A. M. Aguilera, M. Escabias, C. Preda, and G. Saporta, "Using basis expansions for estimating functional pls regression: Applications with chemometric data," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 2, pp. 289-305, 2010.
[26] M. C. Aguilera-Morillo, A. M. Aguilera, M. Escabias, and M. J. Valderrama, "Penalized spline approaches for functional logit regression," *Test*, vol. 22, no. 2, pp. 251-277, 2013.
[27] K. Benhenni, F. Ferraty, M. Rachdi and P. Vieu, "Local smoothing regression with functional data," *Computational Statistics*, vol. 22, no. 3, pp. 353-369, 2007.
[28] A. Baíllo and A. Grané, "Local linear regression for functional predictor and scalar response," *Journal of Multivariate Analysis*, vol. 100, no. 1, pp. 102-111, 2009.
[29] J. Barrientos-Marin, F. Ferraty, and P. Vieu, "Locally modelled regression and functional data," *Journal of Nonparametric Statistics*, vol. 22, no. 5, pp. 617-632, 2010.
[30] T. Alexandrov, J. Decker, B. Mertens, A. M. Deelder, R. A. Tollenaar, P. Maass, and H. Thiele, "Biomarker discovery in maldi-tof serum protein profiles using discrete wavelet transformation," *Bioinformatics*, vol. 25, no. 5, pp. 643-649, 2009.
[31] P. De la Mata-Espinosa, J. Bosque-Sendra, R. Bro, and L. Cuadros-Rodríguez, "Olive oil quantification of edible vegetable oil blends using triacylglycerols chromatographic fingerprints and chemometric tools," *Talanta*, vol. 85, no. 1, pp. 177-182, 2011.
[32] P. De la Mata-Espinosa, J. Bosque-Sendra, R. Bro, and L. Cuadros-Rodriguez, "Discriminating olive and non-olive oils using hplc-cad and chemometrics," *Analytical and bioanalytical chemistry*, vol. 399, no. 6, pp. 2083-2092, 2011.
[33] K. Chen, S. Guo, Y. Lin and Z. Ying, Least absolute relative error estimation. *J. Am. Statist. Assoc.*, vol. 105, 1104–1112, 2010.
[34] J. Demongeot, M. Ghassani, I. Ouassou, C. Taramasco and M. Rachdi (2013). Archimedean copula and contagion modelling in epidemiology Networks and Heterogeneous Media, American Institute of Mathematical Science (AIMS), Vol. 08, Issue 01, Pages 149-170.
[35] J. Demongeot, A. Laksaci, A. Naceri and M. Rachdi (2017). Local linear regression modelization when

all variables are curves Statistics & Probability Letters, Vol. 121, Pages 37-44.

[36] J. Demongeot, A. Laksaci, F. Madani and M. Rachdi (2013). Functional data: local linear estimation of the conditional density and its application Statistics, Vol. 47, Issue 01, Pages 26-44.

[37] F. Ferraty and P. Vieu, *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media, 2006.

[38] A. Laksaci, F. Madani and M. Rachdi (2013). Kernel conditional density estimation when the regressor is valued in a semi-metric space. Communications in Statistics-Theory and Methods, 2013, vol. 42, no. 19, Pages 3544-3570.

[39] P. McCullagh, J.A. Nelder, and P. McCullagh, *Generalized linear models*, vol. 2. Chapman and Hall London, 1989.

[40] M. Rachdi, A. Laksaci, A.M. Almanjahie and Z.C. Elmezouar (2020). FDA: Theoretical and practical efficiency of the local linear estimation based on the kNN smoothing of the conditional distribution when there are missing data. *Journal of Statistical Computation and Simulation*, Vol. 90, Issue 8, Pages 1479-1495.

[41] M. Rachdi and P. Vieu, "Nonparametric regression for functional data: automatic smoothing parameter selection," *Journal of Statistical Planning and Inference*, vol. 137, no. 9, pp. 2784-2801, 2007.

[42] J.O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.