Computational Statistics (2007) 22:223–235
DOI 10.1007/s00180-007-0041-4

ORIGINAL PAPER

PLS classification of functional data

Cristian Preda · Gilbert Saporta · Caroline Lévéder

Published online: 23 February 2007

Abstract Partial least squares (PLS) approach is proposed for linear discriminant analysis (LDA) when predictors are data of functional type (curves). Based on the equivalence between LDA and the multiple linear regression (binary response) and LDA and the canonical correlation analysis (more than two groups), the PLS regression on functional data is used to estimate the discriminant coefficient functions. A simulation study as well as an application to kneading data compare the PLS model results with those given by other methods.

Keywords PLS regression · Functional data · Linear discriminant analysis

1 Introduction

Statistical methods for data representing functions or curves have received much attention in recent years and classical tools from the finite multivariate

C. Preda (✉)
CERIM - Département de Statistique, Faculté de Médecine,
Université de Lille 2, 1, Place de Verdun, 59045 Lille, France
e-mail: cpreda@univ-lille2.fr

G. Saporta
Chaire de Statistique Appliquée, CEDRIC, CNAM, 292,
Rue Saint Martin, 75141 Paris Cedex 03, France
e-mail: saporta@cnam.fr

C. Lévéder
Danone Vitapole, 128 Route Départementale,
91767 Palaiseau Cedex, France
e-mail: caroline.leveder@danone.com

data analysis are adapted to this kind of data. Examples of functional data can be found in several application domains such as medicine, economics, chemometrics and many others (for an overview, see Ramsay and Silverman 2002). A well accepted model for functional data is to consider it as paths of a stochastic process $\mathbf{X} = \{X_t\}_{t\in T}$ taking values in a Hilbert space of functions on some set T. For example, for $T \in \mathbb{R}_+$, a second order stochastic process $\mathbf{X} = \{X_t\}_{t\in[0,T]}$ L_2- continuous with sample paths in $L_2([0, T])$ can be used as model for describing the behavior of some quantitative parameter associated to a process observed on a time interval of length T. A major interest when dealing with functional data is to develop regression models, in particular, linear models. In this case, the multicolinearity feature of the predictor (the X_t's are highly correlated, the covariance operator being, in general, degenerated) produces inconsistency of the classical model estimation. The dimension of the space the observations belong (in general, infinite) with respect to the sample size represents also a difficulty for fitting the model. Depending on the nature of the response variable, several solutions are proposed in literature. Thus, generalized linear regression models are developed in James (2002) and more recently by Cardot and Sarda (2005) and Müler and StadtMüler (2005). Different linear approaches based on decomposition of the underlying stochastic process are proposed : principal component regression (Aguilera et al. 1997; Cardot et al. 1999), partial least squares regression (PLS)(Preda and Saporta 2002). Ferraty and Vieu (2004) and Preda (2007) propose non-parametric models for regression on functional data using classical kernel estimators both for scalar and categorical response.

In this paper we are interested in linear discrimination analysis (LDA) when the predictor $\mathbf{X}$ is data of functional type (generally, curves or real functions of time) and the response is a categorical variable Y defining K groups, $K \geq 2$. As an extension of the classical multivariate approach, the aim of linear discriminant analysis for functional data is to find linear combinations $\Phi(\mathbf{X}) = \int_0^T X_t\beta(t)dt$, $\beta \in L_2([0, T])$ such that the between class variance is maximized with respect to the total variance, i.e.,

$$\max_{\beta\in L_2[0,T]} \frac{\mathbb{V}(\mathbb{E}(\Phi(X)|Y))}{\mathbb{V}(\Phi(X))}. \tag{1}$$

Due to infinite dimension of the predictor, the estimation of β is in general an ill–posed problem. For $K = 2$ is well known that the optimization problem (1) is equivalent to find the regression coefficients of the linear regression of Y (after a convenient encoding) on the stochastic process $\mathbf{X}$. Solutions based on regularization (principal components, for example) and projection (spline basis) techniques have been developed. A linear discriminant analysis model for functional data is proposed by James and Hastie (2001) providing also solutions to the problem of sparse data. Nonparametric approaches are proposed in Biau et al. (2005) and Ferraty and Vieu (2003) for analyzing spectrometric data. Logistic regression using projection methods (splines), respectively the principal component regularization technique is developed in Escabias et al. (2005), respectively in Escabias et al. (2004) with applications to environmental

problems. Ratcliffe et al. (2002) develop functional regression models for foetal heart data in order to predict the probability of high risk birth outcome.

The connection between PLS and LDA is presented in Barker and Rayens (2003) for the finite dimensional, PLS being defined under orthogonal constraints for the vectors giving the PLS scores. It is shown that PLS is to be preferred over PCA when discrimination is the goal and dimension reduction is needed. That is the case for functional data. Our PLS approach is different in that sense we look for uncorrelated PLS components (scores) instead the orthogonality of coefficient functions giving the PLS scores.

We propose to use the PLS regression approach developed in Preda and Saporta (2002) in order to perform LDA on functional data. We derive elements from the structure of the predictor and the response variable (PLS components), which allow to estimate the discriminant coefficient functions and to compute discriminant scores. In Sect. 2 we recall some results on PLS regression on functional data and show how to use it for the discrimination problem. A simulation study as well as an application to kneading data are presented in Sect. 3.

2 PLS regression and LDA for functional data

Let us suppose the functional data is represented by sample paths of a stochastic process $\mathbf{X} = \{X_t\}_{t\in[0,T]}$, $T > 0$. We assume that $\mathbf{X}$ is a second order stochastic process L_2–continuous with sample paths in $L_2[0, T]$ and $\mathbb{E}(X_t) = 0, \forall t \in [0, T]$. These hypothesis ensure, for example, the covariance of $\mathbf{X}$ is a Hilbert–Schmidt operator, and thus, linear analysis of the process is possible. Also, they are not too restrictive from practical point of view, the results presented being still valid if the process is L_2–continuous except a finite set of time points (see for details Saporta 1981).

2.1 PLS regression

Let $\mathbf{Y} = (Y_1, \ldots, Y_p)$ be a real random vector defined on the same probability space as $\mathbf{X}$. It is well known that the estimation of the linear model $\mathbf{Y} = \langle \beta, \mathbf{X} \rangle + \varepsilon$ under least squares criterion is an ill-posed problem in the context of functional data (Cardot et al. 1999).

Dimension reduction is one solution to this problem and principal component regression (PCR) is a typical choice (Aguilera et al. 1997). Replacing the least squares criterion with that of maximal covariance between $\mathbf{X}$ and $\mathbf{Y}$, the PLS regression offers a good alternative to PCR (Preda and Saporta 2002).

The basic idea of PLS approach is to construct a set of uncorrelated random variables $\{t_i\}_{i\geq 1}$ (PLS components) in the linear space spanned by $\mathbf{X}$, taking into account the correlation between $\mathbf{Y}$ and $\mathbf{X}$. We recall here the main results.

Let $\mathbf{W}^X$, respectively $\mathbf{W}^Y$, be the Escoufier's operators (Escoufier 1970) associated to $\mathbf{X}$, respectively to $\mathbf{Y}$, defined by:

$$\mathbf{W}^X Z = \int_0^T E(X_t Z) X_t dt, \quad \mathbf{W}^Y Z = \sum_{i=1}^p E(Y_i Z) Y_i, \quad \forall Z \ r.v.$$

Notice that the eigenvectors of $\mathbf{W}^X$, respectively $\mathbf{W}^Y$, are the principal components of $\mathbf{X}$, respectively of $\mathbf{Y}$. From practical point of view, if $X_1, \ldots, X_n$ is a sample of $\mathbf{X}$ then, an estimator of $\mathbf{W}^X$ is the matrix $\hat{\mathbf{W}}^X$ of size $n \times n$ with entries $\hat{w}_{i,j} = \langle X_i, X_j \rangle_{L_2([0,T])}$.

Proposition 1 (Tucker criterion)

$$\max_{\substack{w,c \\ w \in L_2([0,T]), \|w\|=1 \\ c \in R^p, \|c\|=1}} Cov^2 \left(\int_0^T X_t w(t) dt, \sum_{i=1}^p c_i Y_i \right) \tag{2}$$

is reached for w, respectively c, the eigenvectors associated to the largest eigenvalue of $\mathbf{U}_X = \mathbf{C}_{XY} \circ \mathbf{C}_{YX}$, *respectively of* $\mathbf{U}_Y = \mathbf{C}_{YX} \circ \mathbf{C}_{XY}$, *where* $\mathbf{C}_{YX}$ *and* $\mathbf{C}_{XY}$ *are the cross-covariance operators.*

As noted by Frank and Friedman (1993), PLS can be considered as penalized canonical correlation analysis with penalties provided by PCA in $\mathbf{X}$, respectively in $\mathbf{Y}$.

The random variable $t = \int_0^T X_t w(t) dt$ is called PLS component. A simple way to find t is given by the following proposition.

Proposition 2 *t is the eigenvector associated to the largest eigenvalue of* $\mathbf{W}^X \mathbf{W}^Y$.

PLS iteration The PLS regression is an iterative method.

Let $\mathbf{X}_0 = \mathbf{X}$ and $\mathbf{Y}_0 = \mathbf{Y}$. For $q \geq 1$ we define the qth PLS component, t_q, by the eigenvector associated to the largest eigenvalue of $\mathbf{W}^X_{q-1} \mathbf{W}^Y_{q-1}$, where $\mathbf{W}^X_{q-1}$, respectively $\mathbf{W}^Y_{q-1}$, are the Escoufier's operators associated to $\mathbf{X}_{q-1}$, respectively to $\mathbf{Y}_{q-1}$. The PLS step is completed by the ordinary linear regression of $\mathbf{X}_{q-1}$ and $\mathbf{Y}_{q-1}$ on t_q. Let $\mathbf{X}_q$, and $\mathbf{Y}_q$ be the random variables which represent the residual of these regressions : for each $t \in [0, T]$ and $i \in \{1 \ldots p\}$, $X_{q,t} = X_{q-1,t} - p_q(t) t_q$ and $Y_{q,i} = Y_{q-1,i} - c_{q,i} t_q$.

Then,

Proposition 3 *For each* $q \geq 1$:

(a) $\{t_i\}_{i=1,\ldots q}$ *forms an orthogonal system in* $L_2(X)$,

(b) $Y_i = c_{1,i}t_1 + c_{2,i}t_2 + \cdots + c_{q,i}t_q + Y_{q,i}, \quad i = 1, \ldots, p,$
(c) $X_t = p_1(t)t_1 + p_2(t)t_2 + \cdots + p_q(t)t_q + X_{q,t}, \quad t \in [0, T].$

Thus, the PLS approximation of $\mathbf{Y}$ by $\mathbf{X}$ at step q, $q \geq 1$, is given by:

$$\begin{aligned} \hat{Y}_i &= c_{1,i}t_1 + \cdots + c_{q,i}t_q, \;\; c_{i,j} \in \mathbb{R}^p, \; i = 1, \ldots, p, \; j = 1, \ldots, q, \\ &= \int_0^T \beta_{i,PLS}(t) X_t dt, \quad \beta_{i,PLS} \in L_2([0, T]). \end{aligned} \tag{3}$$

In practice, the number of PLS components to be considered for regression is given by cross validation. For a training sample of size n, let $PRESS(q)$ be the prediction error sum of squares using q components, defined by $PRESS(q) = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_{(-i)}\right)^2$ and $RSS(q)$ be the residual sum of squares defined by $RSS(q) = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$. Then, the PLS component t_q, $q \geq 2$, is retained for regression, if $PRESS(q) \leq (1 - \alpha)RSS(q - 1)$, for some $\alpha \in (0, 1)$. In our applications we choose $\alpha = 0.05$ (see for details Tenenhaus 2002).

Remark 1 (Computational aspects)

(i) For each step q, $q \geq 2$, the cross validation requires to repeat n times the PLS procedure with samples of size $n - 1$.
(ii) Provided that the process $\mathbf{X}$ is observed for each $t \in [0, T]$, numerical approximation is necessary in order to compute the PLS components. A quadrature method is in general sufficient in order to compute the entries of the matrix $\hat{W}$. Discretizing the time interval, one obtains also an approximation for β_{PLS}.
(iii) If $\mathbf{X}$ is observed only in a finite number of time points, then it is necessary to approximate, for each observation, the true form of the corresponding curve. The interpolation procedure is appropriate if data are observed without error. Otherwise, one common method is to choose a finite dimensional basis of functions and to project each observation onto this basis. The coefficients of basis functions could be computed using, for example, a penalized version of the least squares criterion. See for example Ramsay and Silverman (1997), Chap. 3.

The connection between linear regression and LDA is well known in finite dimension. For binary response, the coefficient vector giving the discriminant score is obtained by linear regression after a convenient encoding of the response. In the next section we show that is also true for functional data, but LDA can not be applied directly because of the infinite dimension of the predictor. Then, in a natural way, we propose the PLS regression in order to estimate the discriminant coefficient function.

2.2 LDA and linear regression

LDA can not be directly applied to infinite dimensional data such the functional data. A functional linear discriminant analysis (FLDA) model is developed by James and Hastie (2001) for irregularly sampled curves using natural cubic spline basis to represent data and the corresponding coefficients for classification.

Here we are interested to the optimization problem stated in (1), which is the formulation given by Fisher in 1936 for LDA (Fisher 1936).

Let us consider for instance the case of binary response, $\mathbf{Y} \in \{0, 1\}$.

Denote by $p_0 = \mathbb{P}(\mathbf{Y} = 0), p_1 = 1 - p_0 = \mathbb{P}(\mathbf{Y} = 1)$ and $\mu_0(t) = \mathbb{E}(X_t|\mathbf{Y} = 0)$, $\mu_1(t) = \mathbb{E}(X_t|\mathbf{Y} = 1), t \in [0, T]$.

Since $\mathbb{E}(X_t) = 0$, it follows that $p_0\mu_0(t) + p_1\mu_1(t) = 0, \forall t \in [0, T]$. Let also $\mathbf{C}$ be the covariance operator associated to the process $\mathbf{X}$ defined on $L_2[0, T]$ by

$$f \overset{\mathbf{C}}{\longmapsto} g, \quad g(t) = \int_0^T C(t, s)f(s)ds,$$

where $C(t, s) = \mathbb{E}(X_tX_s)$. Let $\mathbf{B}$ be the operator on $L_2[0, T]$ defined by

$$f \overset{\mathbf{B}}{\longmapsto} g, \quad g(t) = \int_0^T B(t, s)f(s)ds,$$

where $B(t, s) = p_0\mu_0(t)\mu_0(s)+p_1\mu_1(s)\mu_1(t) = p_0p_1(\mu_0(t)-\mu_1(t))(\mu_0(s)-\mu_1(s))$. Denoting by $\phi = \sqrt{p_0p_1}(\mu_0 - \mu_1)$, it follows that

$$\mathbf{B} = \phi \otimes \phi,$$

where $\phi \otimes \phi(g) = \phi\langle\phi, g\rangle_{L_2[0,1]}, g \in L_2[0, T]$.

As in the finite dimensional setting, the discriminant coefficient function, $\beta \in L_2[0, T]$, which satisfies the criterion given in (1), corresponds to the largest λ, $\lambda \in \mathbb{R}$, such that

$$\mathbf{B}\beta = \lambda\mathbf{C}\beta, \quad \text{with } \langle\beta, \mathbf{C}\beta\rangle_{L_2[0,T]} = 1. \tag{4}$$

Let us recode $\mathbf{Y}$ by $0 \rightsquigarrow \sqrt{\frac{p_1}{p_0}}$ and $1 \rightsquigarrow -\sqrt{\frac{p_0}{p_1}}$. If β is a solution of (4) then $\lambda = \langle\phi, \beta\rangle^2_{L_2[0,T]}$ and β is also solution of the equation

$$\mathbb{E}(YZ_t) = \int_0^T \mathbb{E}(Z_tZ_s)\beta(s)ds, \tag{5}$$

where $Z_t = \sqrt{\lambda}X_t$, $t \in [0, T]$. The Wiener-Hopf equation (5) is equivalent to the least squares criterion for the linear regression of $\mathbf{Y}$ on the process

$\mathbf{Z} = \{Z_t\}_{t\in[0,T]}$. Thus, the discriminant coefficient function β is, up to a constant, the regression coefficient function of the linear regression of $\mathbf{Y}$ on $\mathbf{X}$. Conversely, any solution of (5), up to a constant, is also solution for (1). But Eq. (5) has an unique solution in $L_2([0,T)]$ if and only if convergence of series implying eigenvalues and eigenvectors of the covariance operator $\mathbf{C}$ is ensured (see Picard theorem in Saporta 1981, p. 158). These conditions are rarely satisfied, the problem stated in (5), and so in (1), being, in general, ill-posed.

This connection between LDA and linear regression shows that, in general, one cannot write the discriminant score $\Phi(\mathbf{X})$ as linear combination of variables $\mathbf{X}$, $\Phi(\mathbf{X}) = \int_0^T X_t\beta(t)dt$, with $\beta \in L_2([0,T])$. The discriminant score is the orthogonal projection of $\mathbf{Y}$ (after appropriate encoding) on the linear space spanned by $\{X_t,\ t \in [0,T]\}$. Then, regularized techniques for linear regression can be used in order to derive an approximation for the discriminant score. If the goal of LDA is the classification, the coefficient regression function so obtained allows to predict for new observations the discriminant score.

If the response has more than two categories ($K > 2$) then LDA is equivalent to canonical correlation analysis (Saporta 1981). However, as for the binary response case, the canonical factor for $\mathbf{X}$ (which is also the discriminant coefficient function) is solution of an equation of type (5). Considered as a penalized canonical correlation analysis, PLS approach provides a regularization method for LDA by using the PLS components as predictors.

Using the same arguments as in Barker and Rayens (2003) (Sect. 3.2) in the context of functional data, one can consider that the PLS approach is a natural alternative to PCA, and is to be preferred especially when dimension reduction is necessary and the within–groups variability dominates the among–groups variability. In addition, de Jong (1993) shows that PLS fits closer than the linear regression on principal components and thus, at least for $K = 2$, it is clear that PLS discriminant model is to be preferred to the principal components one.

In view of the above considerations, we propose the following methodology for LDA when PLS regularization is used. The obtained model will be quoted by PLS_FLDA.

PLS classification methodology for functional data Let $\mathbf{Y}$ be a categorical response taking values in $\{1,\ldots,K\}$.

Case 1: $K = 2$. The discriminant function β is the coefficient function of the linear regression of $\mathbf{Y}$ on $\mathbf{X}$, where $\mathbf{Y}$ is encoded by $0 \rightsquigarrow \sqrt{\frac{p_1}{p_0}}$ and $1 \rightsquigarrow -\sqrt{\frac{p_0}{p_1}}$, with (p_0,p_1) the probability distribution of $\mathbf{Y}$. The PLS regression of $\mathbf{Y}$ on $\mathbf{X}$ provides an approximation for the discriminant variable (score) given by $\Phi_{PLS}(\mathbf{X}) = \alpha + \int_0^T X_t\beta_{PLS}(t)dt$, where $\alpha = -\int_0^T \beta_{PLS}(t)\mu(t)dt$, $\mu(t) = \mathbb{E}(X_t), t \in [0,T]$.

Case 2: $K > 2$. Let $\{Y_i\}_{i=1,\ldots,K-1}$ be the dummy variables associated to $\mathbf{Y}$, $Y_i \in \{0,1\}$. We propose as functional discriminant analysis PLS model for $\mathbf{Y}$ and X, the model obtained by performing the classical LDA of $\mathbf{Y}$ and the set of PLS components $\{t_j\}_{j=1,\ldots,q}$ obtained from the PLS regression of the vector $\{Y_1,\ldots,Y_{K-1}\}$ on $\mathbf{X}$. One obtains, for each

category i of $\mathbf{Y}$, the coefficient discriminant function β^i_{PLS} which allow to compute the associated score, $\Phi^i_{PLS}(\mathbf{X}) = \alpha_i + \int_0^T X_t \beta^i_{PLS}(t)dt$, with $\alpha_i \in \mathbb{R}, i = 1, \ldots, K$.

Therefore, given a new observation of $\mathbf{X}$, prediction for $\mathbf{Y}$ is made in the classical way. The predictive capacity of the method is then measured using either the misclassification rate or the area under the ROC curve (for $K = 2$).

3 Numerical applications

We present a simulation study as well as an application to kneading data and compare the results of the functional discriminant PLS approach with those given by other methods. Thus, we quoted by

- PC_FLDA : the model obtained by regularization using principal components. The principal components in the model are ordered by their explained variance rate from the most explicative to the least one. Their number is determined by cross validation procedure, as presented in the Remark of Sect. 2.1.
- K-NN(k) : the K-NN rule classification with k neighbors and L_1 distance. The number of neighbors in the K-NN procedure is such that it ensures minimum error rate classification among all values from a predefined set (Lévéder et al. 2004).
- Gaussian(σ) : the non-parametric model developed in Preda (2007) using reproducing kernel Hilbert space (RKHS) methods with Gaussian kernel of parameter σ.
- LDA : the model built using the predictors given by a finite discretization of time interval and a stepwise procedure for predictor's selection (SAS procedure).

For PLS_FLDA, PC_FLDA and RKHS models we developed software implemented in *C* language.

3.1 Simulation study

The simulated data we consider correspond to a binary response for which the predictor has the following form :

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

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

where U is a r.v. uniformly distributed on $[0, 1]$, $\varepsilon(t)$ are uncorrelated standard normal r.v.'s and $h_1(t) = \max\{6-|t-11|, 0\}, h_2(t) = h_1(t-4)$ and $h_3(t) = h_1(t+4)$. As in Ferraty and Vieu (2003), we consider the observed predictor is a discretized curve with 101 equidistant points $\{t = 1, 1.2, 1.4, \ldots, 21\}$. Figure 1 displays a sample of 100 simulated curves for each class.

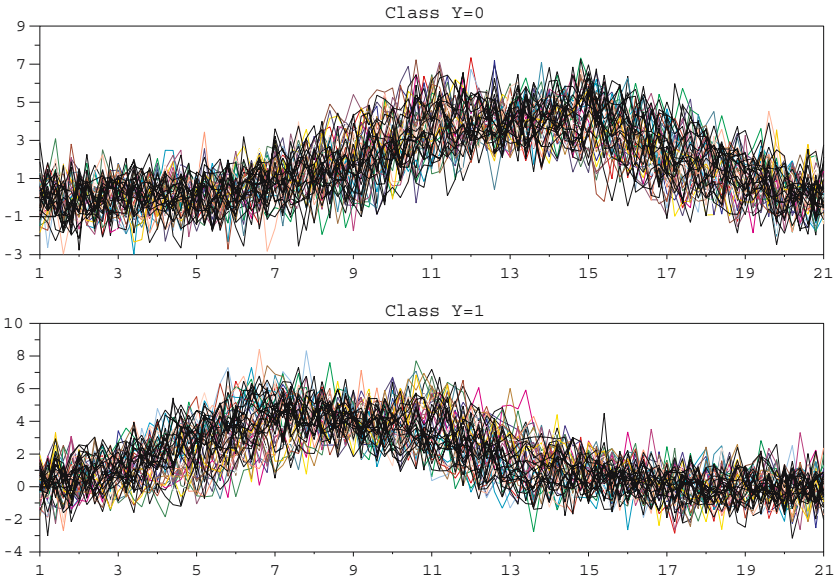

Fig. 1 Sample of 100 curves for each class

Table 1 Error rate averaged over 100 test samples

Model	PLS_FLDA	K-NN(15)	PC_FLDA	Gaussian(6)	LDA
Error rate	0.0286	0.0265	0.0332	0.0255	0.0813

As in Preda (2007), we consider 100 simulated samples of size 1000, with 500 observations in each class. Each sample is randomly divided into a training sample of size 800 and a test sample of size 200, each class having the same number of observations in both samples.

The principal components of the process $\{X_t\}_{t\in[1,21]}$, as well as the PLS components are computed using linear interpolation and the trapezoidal integration method.

Table 1 presents the error rates averaged over the 100 test samples given by different models.

All models give good results with respect to the classification error rate criterion. PLS approach gives the best results among the considered linear models.

3.2 Application to kneading data

PLS approach is applied to predict the quality of cookies from curves representing the resistance (density) of dough observed during the kneading process. For a given flour, the resistance of dough is recorded during the first

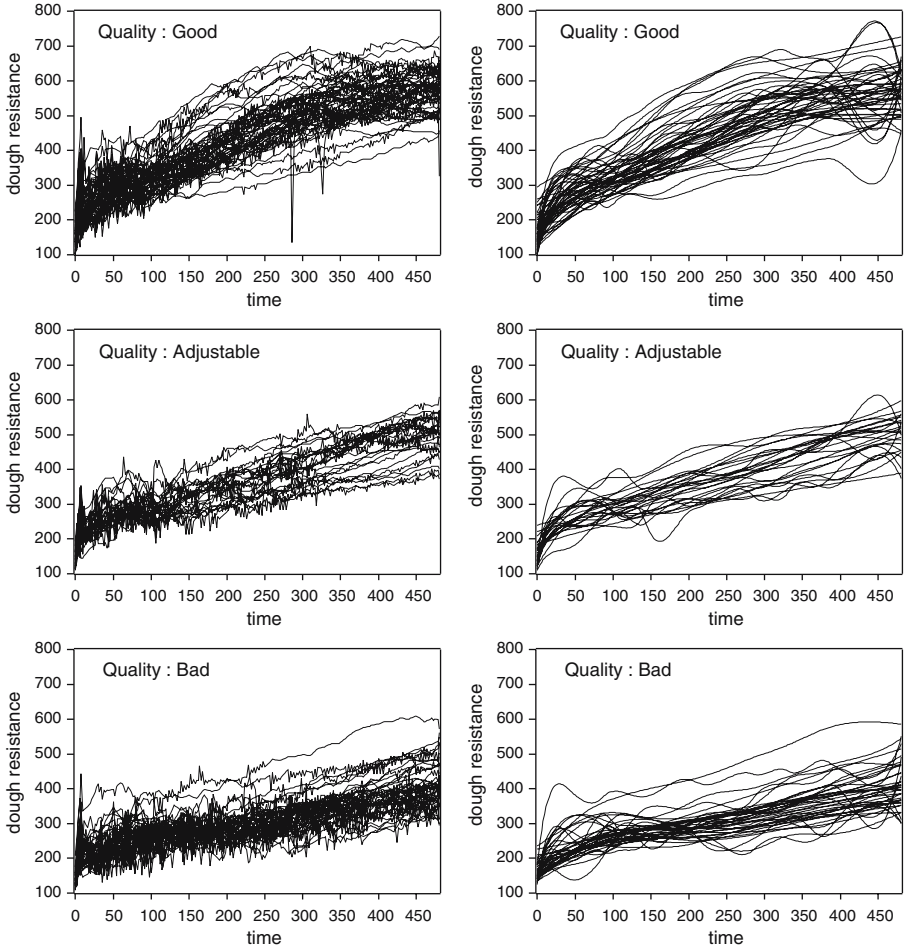

Fig. 2 Kneading data : 115 flours observed during 480 s. *Left* observed data. *Right* smoothed data using cubic B-splines

480 s of the kneading process. For 115 different flours we have 115 curves which can be considered as sample paths of a L_2-continuous stochastic process, $\mathbf{X} = \{X_t, t \in [0, 480]\}$. Each curve is observed in 240 equispaced time points of the interval time $[0, 480]$. Thus, a kneading curve is represented by the set of 241 points $\{(t_i, X_{t_i}), i = 0, \ldots, 240\}$.

After kneading, the dough is processed to obtain cookies. For each flour we have the quality ($\mathbf{Y}$) of cookies which can be *Good, Adjustable* and *Bad*. Our sample contains 50 observations for $\mathbf{Y} = Good$, 25 for $\mathbf{Y} = Adjustable$ and 40 for $\mathbf{Y} = Bad$ (Fig. 2). Due to measuring errors, each curve is smoothed using cubic B-spline functions with the following 16 knots (Lévéder et al. 2004) : {10, 42, 84, 88, 108, 134, 148, 200, 216, 284, 286, 328, 334, 380, 388, 478} (right side in Fig. 2).

Table 2 Error rate averaged over 100 test samples

Model	PLS_FLDA	K-NN(13)	PC_FLDA	Gaussian(6)	LDA
Error rate	0.112	0.103	0.142	0.108	0.154

Table 3 Misclassified rate for three categories : average over 100 test samples

Model	PLS_FLDA	K-NN(13)	PC_FLDA	Gaussian(6)	LDA
Error rate	0.258	0.245	0.262	0.247	0.282

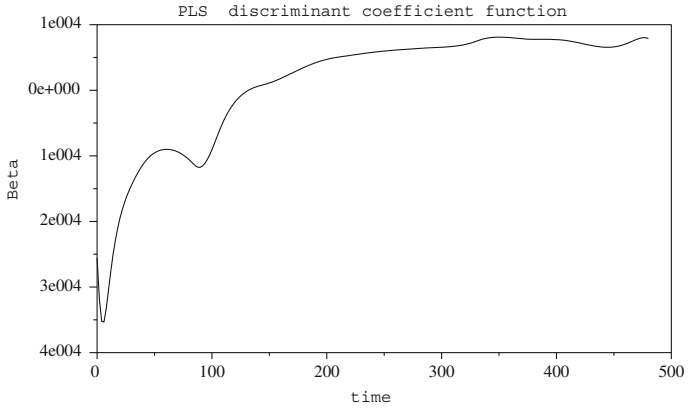

Fig. 3 Discriminant coefficient function $\hat{\beta}_{PLS}$ given by PLS approach

Let us consider $\mathbf{Y} \in \{Good, Bad\}$. The sample of 90 flours is randomly divided into a learning sample of size 60 and a test sample of size 30. In the test sample the two classes have the same number of observations. The PLS and the principal components are computed from smoothed data using the trapezoidal integration method with the 241 equidistant time points.

Table 2 presents the error classification rates averaged over 100 test samples.

It is useful to study the set of observations for which $\mathbf{Y} = Adjustable$: with a small effort, some of these flours could be *adjusted* to become *Good*. Therefore, for a flour with quality *Adjustable* is important to decide if it is closer to *Good* or to *Bad* quality. For this purpose, we consider the set of *Adjustable* flours as test sample and predict for each one the group membership, $\mathbf{Y} \in \{Good, Bad\}$, using the discriminant coefficient function (Fig. 3) given by the PLS approach on the 90 flours. The discriminant score is $\Phi(\mathbf{X}) = -1.46 + \int_0^{480} X_t \hat{\beta}_{PLS}(t)dt$. For a given x, if $\Phi(x) > 0$ then $\hat{Y} = Good$, else $\hat{Y} = Bad$.

One obtains for the 25 *Adjustable* flours, 12 flours for which the prediction gives them closer to class *Good*.

For $\mathbf{Y} \in \{Good, Adjustable, Bad\}$, the results given by the multivariate PLS discrimination model (Sect. 2.2, case 2) are presented in Table 3. The comparison

criterion is the averaged error classification rate over 100 test samples of size 35.

4 Conclusion

PLS regression is used for linear discriminant analysis when the predictors are curves or functions. PLS approach is a simple and interesting alternative to classical linear methods based on principal components, giving in general better results. The proposed PLS methodology is studied on simulated and real data.

Acknowledgments The authors like to thank the Editor and referees for many constructive suggestions. We are grateful to Group Danone Vitapole, Centre de Recherche Daniel Carasso, Paris, for providing us with data for testing our methods.

References

Aguilera AM, Ocaña F, Valderama MJ (1997) An approximated principal component prediction model for continuous-time stochastic process. Appl Stochast Models Data Anal 13:61–72
Barker M, Rayens W (2003) Partial least squares for discrimination. J Chemomet 17:166–173
Biau G, Bunea F, Wegkamp M (2005) Function classification in Hilbert spaces. IEEE Trans Inform Theory 51:2162–2172
Cardot H, Sarda P (2005) Estimation in generalized linear models for functional data via penalized likelihood. J Multivariate Anal 92:24–41
Cardot H, Ferraty F, Sarda P (1999) Functional linear model. Statist Probab Lett 45:11–22
Escabias M, Aguilera AM, Valderrama MJ (2004) Principal component estimation of functional logistic regression: discussion of two different approches. J Nonparamet Stat 3–4:365–385
Escabias M, Aguilera AM, Valderrama MJ (2005) Modeling environmental data by functional principal component logistic regression. Environmetrics 16:95–107
Escoufier Y (1970) Echantillonage dans une population de variables aléatoires réelles. Publications de l'Institut de Statistique de l'Université de Paris 19(4):1–47
Ferraty F, Vieu P (2003) Curves discrimination: a nonparametric approach. Comput Stat Data Anal 44:161–173
Ferraty F, Vieu P (2004) Nonparametric models for functional data with application in regression, time series prediction and curve discrimination. J Nonparamet Stat 16(1–2):111–125
Fisher RA (1936) The use of multiple measurement in taxonomic problems. Ann Eugen 7:179–188
Frank I, Friedman J (1993) A statistical view of some chemometrics regression tools. Technometrics 35:109–148
James GM (2002) Generalized linear models with functional predictors. J R Stat Soc Ser B 64(3):411–432
James GM, Hastie TJ (2001) Functional discrimnant analysis for irregularly sampled curves. J R Stat Soc Ser B 63(3):533–550
de Jong S (1993) Pls fits closer than pcr. J Chemomet 7:551–557
Lévéder C, Abraham C, Cornillon PA, Matzner-Lober E, Molinari N (2004) Discrimination de courbes de pétrissage. Chimiométrie 2004, Paris, pp 37–43
Müler H-G, StadtMüler U (2005) Generalized functional linear models. Ann Stat 33(2):774–805
Preda C (2007) Regression models for functional data by reproducing kernel Hilbert spaces methods. J Stat Plan Infer 137:829–840
Preda C, Saporta G (2002) Régression PLS sur un processus stochastique. Revue de Statistique Appliquée L(2):27–45
Ramsay JO, Silverman BW (1997) Functional data analysis. Springer, New York
Ramsay JO, Silverman BW (2002) Applied functional data analysis:methods and case studies. Springer, Heidelberg

Ratcliffe SJ, Leader LR, Heller GZ (2002) Functional data analysis with application to periodically stimulated foetal heart rate data. ii : functional logistic regression. Stat Med 21:1115–1127
Saporta G (1981) Méthodes exploratoires d'analyse de données temporelles. Cahiers du B.U.R.O, Université Pierre et Marie Curie, Paris
Tenenhaus M (2002) La régression PLS. Théorie et pratique. Editions Technip