

Comments on: Probability enhanced effective dimension reduction for classifying sparse functional data

Ana M. Aguilera¹

Published online: 25 January 2016
© Sociedad de Estadística e Investigación Operativa 2016

First of all, the authors deserve to be congratulated for their interesting and original contribution on functional data classification. The main goal is classifying a set of sparsely observed curves in two groups defined by a binary response variable. Classification of curves is a major problem that has been widely studied from different points of view on recent research in FDA. However, the classification of sparse functional data has been scarcely studied and therefore the current work is even more valuable. In what follows some additional references and comments that can be useful for completing the paper by Yao, Wu, and Zou are provided.

In this paper, the authors consider the projection of the functional data onto the most effective directions associated with a functional index model. A new functional cumulative slicing approach, based on ranking the probabilities obtained by weighted support vector machine, is proposed for the estimation of the direction functions with a binary response and sparse functional data. Their approach is related with one of the main lines of research on binary classification of a set of curves. This is based on dimension reduction of the infinite-dimensional functional space and application of a commonly used classifier (LDA, logit, etc.) to a reduced set of uncorrelated variables obtained by projection onto the resultant space of directions.

This comment refers to the invited paper available at: doi:[10.1007/s11749-015-0470-2](https://doi.org/10.1007/s11749-015-0470-2).

This research was supported by Projects MTM2013-47929-P from Dirección General de Investigación Científica y Técnica del Ministerio de Economía y Competitividad, Spain, and FQM-08068 from Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía, Spain.

✉ Ana M. Aguilera
aaguiler@ugr.es

¹ Departamento de Estadística e I.O. Facultad de Ciencias, Universidad de Granada, Campus de Fuentenueva s/n, 18071 Granada, Spain

The most used methodology for dimension reduction with functional data is undoubtedly functional principal component analysis (FPCA) introduced by [Devillette \(1974\)](#) in a really extensive and valuable work (almost one hundred pages) with a different name: harmonic stochastic analysis for stochastic processes with squared integrable functions. Asymptotic theory and statistical inference for PCA of a random variable with values in a separable Hilbert space were later studied in [Dauxois et al. \(1982\)](#). In the theoretical context given by Hilbert valued random variables, equivalences between PCA of the original functional data with respect to a well-suited geometry (inner product) and PCA of transformed data with respect to a given geometry were studied by [Ocaña et al. \(1999\)](#) (filtering PCA). In many applications, FPCA has to be estimated from discrete observations of sample curves that could be even irregularly sampled, noisy, and sparse. FPCA estimation is usually based on the approximation of sample curves in terms of basis functions (splines, wavelets, trigonometric functions, etc.) The algorithm for computing FPCA from basis expansions of sample curves is based on the classic multivariate PCA of certain random vector given by a transformation of the vector of basis coefficients. In fact, this algorithm is derived from the analysis of the effects of modifying the norm in the space of coordinates ([Ocaña et al. 2007](#)).

The main problem when using FPCA for regression or classification purpose is that it does not take into account the relationship between the predictor functional variable and the response variable. In fact, it usually happens that the most explicative principal components (pc's) are not the best predictors. The difficulties related with the choice of pc's used as predictors for regression were discussed in detail by [Saporta \(1981\)](#). In order to get accurate predictions with an effective dimension reduction, a model selection procedure based more on the predictive ability of the pc's that in their variability order must be considered. See how to solve this problem using stepwise selection ([Escabias et al. 2004](#)), AIC and BIC ([Müller and Stadtmüller 2005](#)) or a mixed criterion based on correlation and explained variance ([Aguilera et al. 1999](#)).

A more advantageous alternative to functional principal component regression or functional classification is partial least squares (PLS). The functional partial least squares regression model was introduced in [Preda and Saporta \(2005\)](#) to solve the problems of high dimension and multicollinearity associated with the scalar-on-function linear model. PLS components are linear functionals of the functional predictor computed by maximizing their squared covariance with the response (Tucker's criterion). In each step, they are obtained by an iterative procedure based on the residuals of the regression of the response and predictor variables on the PLS component computed in the previous step. A detailed study about theoretic properties and explicit formulation of functional PLS was developed in [Delaigle and Hall \(2012a\)](#).

There are some interesting works comparing functional PLS regression (FPLSR) and functional principal component regression (FPCR) ([Reiss and Ogden 2007](#); [Aguilera et al. 2010](#); [Delaigle and Hall 2012a](#)). The main conclusions from them are that FPLSR is superior to FPCR because, although the prediction ability of both approaches is similar, FPLSR requires fewer components to capture the same interactions and provides a much more accurate estimation of the parameter function than FPCR. The smaller dimension of the projection space can be advantageous when computing time is an issue. For example, the PEFCS procedure proposed in this paper performs in the

first step a functional principal component decomposition of the sparsely observed curves via PACE algorithm. The authors suggest to use an initial number of pc's that explain as much as possible of the total variance (nearly 100 %). This previous approximation of the curves has in my opinion two main drawbacks. The first is the computational cost that can have in the majority of cases the computation of a high number of pc's as a previous step to the estimation of EDR directions by PEFCS. The second is related with the pc's discarded in this first approximation that could be highly associated with the binary response. Both problems could be addressed by projection on a subspace of PLS components that explain the majority of the covariance of the predictor and response variables. This involves estimating the PLS model for sparse functional data which could not be a straightforward task.

Functional PLS has been also used for binary classification with different classifiers. For estimating the functional logit model, [Escabias et al. \(2007\)](#) proposed and adaptation of the PLS algorithm for generalized linear models introduced by [Bastien et al. \(2005\)](#). A functional LDA-PLS classification approach based on functional PLS regression of the recoded binary response on the functional predictor was proposed in [Preda et al. \(2007\)](#). On the other hand, a PLS estimation of the centroid-based classifier was considered in [Delaigle and Hall \(2012b\)](#). In these works, the classification results provided by these PLS approaches were compared on extensive simulation studies with the ones given by projecting the original functional data on a reduced set of principal components. Similar to the references on functional regression mentioned above, they conclude that PLS is able to quickly find the relevant components for classification and, as a result has advantages in cases where principal component methods would require a large number of terms.

In the present paper, the authors conclude from data simulated with three different non-linear functional index models that the classification results provided by applying common classifiers (Logit, LDA, QDA, and centroid method) to the data projected onto the estimated EDR directions (PEFCS) are always superior to those based on FPCA, with centroid method and PEFCS having the best classification ability. It is interesting to observe that the differences between the average misclassification errors provided by PEFCS and FPCA with the centroid method are very much bigger than the ones given by the other classifiers. In my opinion, the classification results given by LDA and logistic regression with FPCA are very competitive if we take into account that their computational cost is much lower, the pc's are computed without taking into account the covariance between response and predictor and the simulated models are not propitious for these estimation approaches. Because of this and taking into account the previously mentioned references on functional PLS, I think that a functional PLS estimation approach could improve the results. On the other hand, a good alternative for non-sparse functional data could be a P-spline estimation of functional PCA ([Aguilera-Morillo et al. 2013](#)).

To finish, I would like to thank the editors of *Test* for inviting me to make this discussion and to acknowledge again the authors for this stimulating paper on classification with sparse functional data.

References

- Aguilera AM, Ocaña FA, Valderrama MJ (1999) Forecasting time series by functional PCA. Discussion of several weighted approaches. *Comput Stat* 14:443–467
- Aguilera AM, Escabias M, Preda C, Saporta G (2010) Using basis expansion for estimating functional PLS regression. Applications with chemometric data. *Chemom Intell Lab Syst* 104:289–305
- Aguilera-Morillo MC, Aguilera AM, Escabias M, Valderrama MJ (2013) Penalized spline approaches for functional logit regression. *Test* 22:251–277
- Bastien P, Esposito-Vinzi V, Tenenhaus M (2005) PLS generalised linear regression. *Comput Stat Data Anal* 48(1):17–46
- Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J Multivar Anal* 12:136–154
- Delaigle A, Hall P (2012a) Methodology and theory for partial least squares applied to functional data. *Ann Stat* 40(1):322–352
- Delaigle A, Hall P (2012b) Achieving near perfect classification for functional data. *J R Stat Soc Ser B* 74(2):267–286
- Deville JC (1974) Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE* 15:3–101
- Escabias M, Aguilera AM, Valderrama MJ (2004) Principal component estimation of functional logistic regression: discussion of two different approaches. *J Nonparametr Stat* 16(3–4):365–384
- Escabias M, Aguilera AM, Valderrama MJ (2007) Functional PLS logit regression model. *Comput Stat Data Anal* 51(10):4891–4902
- Müller HG, Stadtmüller (2005) Generalized functional linear models. *Ann Stat* 33(2):774–805
- Ocaña FA, Aguilera AM, Escabias M (2007) Computational considerations in functional principal component analysis. *Comput Stat* 22(3):449–465
- Ocaña FA, Aguilera AM, Valderrama MJ (1999) Functional principal components analysis by choice of norm. *J Multivar Anal* 71:262–276
- Preda C, Saporta G (2005) PLS regression on a stochastic process. *Comput Stat Data Anal* 48:149–158
- Preda C, Saporta G, Lévêder C (2007) PLS classification for functional data. *Comput Stat* 22:223–235
- Reiss PT, Ogden RT (2007) Functional principal component regression and functional partial least squares. *J Am Stat Assoc* 102(479):984–996
- Saporta G (1981) Méthodes exploratoires d'analyse de données temporelles. *Cahiers du B.U.R.O.* 37–38. Universit Pierre et Marie Curie, Paris