

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

G rard Biau^{1,2} · Cl ment Levrard³

Published online: 25 January 2016
  Sociedad de Estadística e Investigaci n Operativa 2016

In this elegant paper, F. Yao, Y. Wu, and J. Zou offer a unified treatment of the problem of classifying sparse functional data via sliced inverse regression (e.g., Li 1991). Such signals are typically encountered in longitudinal studies and various other scientific experiments. In this setting, only a few measurements are available for some, or even all, individuals, and a cumulative slicing approach is proposed by the authors to borrow information across individuals and recover the central subspace.

At first, the authors address the structural issue due to binary classification: since the label Y is discrete, the conditional expectation based on Y can only provide one direction of the central subspace. Therefore, the regression function $p(x) := \mathbb{P}(Y = 1|X = x)$ is preferred as a response variable, in line with the approach of Shin et al. (2014). The function p is estimated by using a weighted support vector machine (SVM) scheme, which is the first step of the probability-enhanced functional cumulative slicing (PEFCS) procedure. In this respect, it is interesting to note that the optimal situation for classification—that is, $p(x) = 0$ or $p(x) = 1$ —is indeed the worst case for this dimension reduction approach, since then only one direction of the central subspace can be recovered.

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).

✉ Cl ment Levrard
levrard@math.univ-paris-diderot.fr
G rard Biau
gerard.biau@upmc.fr

¹ Sorbonne Universit s, UPMC Univ Paris 06, 75005 Paris, France

² Institut universitaire de France, Paris, France

³ Universit  Paris Diderot, Paris, France

Next, the authors address the problem of classifying sparsely observed signals. As pointed out in Section 3, a direct estimation of functional principal components (FPC) scores may lead to poor results in this context. Thus, the principal analysis by conditional expectation (PACE) method, originally presented by Yao et al. (2005), is preferred to estimate the covariance operator and its spectral decomposition, yielding a decomposition of signals in a finite-dimensional space. Finally, the issue of estimating conditional covariances in the sparse setting is addressed via a cumulative slicing strategy, by adapting arguments of Fan and Gijbels (1996) and Yao et al. (2005).

An interesting conclusion of the experimental studies provided in Section 4 of the article is that dimension reduction via central subspace estimation seems to boost the performance of the centroid method of Delaigle and Hall (2012a). The comparison is drawn with respect to FPCA, which plays the role of a benchmark, and with respect to three other classifiers (Linear Discriminant Analysis, Quadratic Discriminant Analysis, and additive Logistic Regression). All procedures are outperformed by the PEFCS + centroid package, at least in the proposed numerical illustrations. The poorer results of FPCA + centroid method show that the directions which capture the more total variation might not be the most useful ones for classification, as already pointed out by Delaigle and Hall (2012a). Thus, the experimental results suggest that the proposed PEFCS algorithm succeed in recovering these useful directions. Nonetheless, it could be of interest to compare the performance of the centroid method based on the reduced data with that of the “raw” centroid method, which could take advantage of the high-dimensional structure, at least if the variance term is neglected (Delaigle and Hall 2012a).

It is also interesting to notice that a preliminary projection, based on PACE, is performed by the PEFCS procedure, prior to the dimension reduction step. Although the authors choose the truncation so that “nearly 100 % of the total variation is explained”, the remaining almost 0 % could contain useful information for classification in some “not-so-pathological” cases. Therefore, a further study might consider choosing the initial projection based on a functional partial least squares basis (PLS; see, e.g., Delaigle and Hall 2012b, or Preda et al. 2007), where priority is given in the functional decomposition to directions that capture the maximal correlation with the response variable. The extension of PLS decomposition to sparse functional data may not be so straightforward. However, combining an initial projection based on PLS with a dimension reduction technique such as the one proposed in the paper may offer more guarantees for classification performance—especially if it is performed through the centroid method. It is likely that if the dimension reduction is carried out only through ranking and thresholding the PLS scores, then the resulting space will be larger than the central subspace. Indeed, the directions that are not in the central subspace but are strongly correlated with directions in this subspace will eventually be selected by such a method. More generally, this remark enlightens a major advantage of the central subspace paradigm over correlation-based methods such as, for instance, the Lasso: correlation between informative directions and non-informative ones does not affect the performance of central subspace recovery, at least theoretically.

Let us finally point out that a drawback of central subspace recovery via inverse regression is the linearity requirement (Assumption 2 in the paper by F. Yao, Y. Wu, and J. Zou), which holds in particular when the distribution of X is elliptic. How-

ever, as shown by [Fukumizu et al. \(2009\)](#), such an assumption may be avoided using a different characterization of the central subspace. The key is to derive a criterion of conditional independence expressed in terms of conditional covariance operators over a Reproducing Kernel Hilbert Space (RKHS) embedding of X . This leads to the construction of the so-called Kernel dimension reduction (KDR) method, which amounts to optimize a contrast function over the Stiefel manifold. This procedure exhibits good performance, even in non-elliptical cases. However, the results of [Fukumizu et al. \(2009\)](#) are derived in a finite-dimensional regression setting. Therefore, a challenging project for the future is to extend the [Fukumizu et al. \(2009\)](#) paradigm to the sparse functional case by following a route similar to the PEFCS strategy.

Thus, to summarize, central subspace recovery is an active and promising research area, which is clearly of interest in high-dimensional settings such as functional data classification. More care is needed when data are sparsely observed and the response variable is binary. This point is remarkably addressed in the paper by F. Yao, Y. Wu, and J. Zou, which we think will stimulate further research in the domain.

References

- Delaigle A, Hall P (2012a) Achieving near perfect classification for functional data. *J R Stat Soc Ser B. Stat Methodol* 74:267–286
- Delaigle A, Hall P (2012b) Methodology and theory for partial least squares applied to functional data. *Ann Stat* 40:322–352
- Fan J, Gijbels I (1996) Local polynomial modelling and its applications, vol 66. Chapman & Hall, London
- Fukumizu K, Bach FR, Jordan MI (2009) Kernel dimension reduction in regression. *Ann Stat* 37:1871–1905
- Li K-C (1991) Sliced inverse regression for dimension reduction. *J Am Stat Assoc* 86:316–342 With discussion and a rejoinder by the author
- Preda C, Saporta G, Lévéder C (2007) PLS classification of functional data. *Comput Stat* 22:223–235
- Shin SJ, Wu Y, Zhang HH, Liu Y (2014) Probability-enhanced sufficient dimension reduction for binary classification. *Biometrics* 70:546–555
- Yao F, Müller H-G, Wang J-L (2005) Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 100:577–590