COMSTA: 6596 Model 3G pp. 1–14 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Sensible functional linear discriminant analysis

Lu-Hung Chen [a], Ci-Ren Jiang [b,*]

[a] Institute of Statistics, National Chung-Hsing University, Taichung, Taiwan

[b] Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 14 October 2017
Received in revised form 9 April 2018
Accepted 14 April 2018
Available online xxxx

Keywords:
Classification
Functional data
Linear discriminant analysis
Longitudinal data
Smoothing

ABSTRACT

Fisher's linear discriminant analysis (LDA) is extended to both densely recorded functional data and sparsely observed longitudinal data for general c-category classification problems. An efficient approach is proposed to identify the optimal LDA projections in addition to managing the noninvertibility issue of the covariance operator emerging from this extension. To tackle the challenge of projecting sparse data to the LDA directions, a conditional expectation technique is employed. The asymptotic properties of the proposed estimators are investigated and asymptotically perfect classification is shown to be achievable in certain circumstances. The performance of this new approach is further demonstrated with both simulated data and real examples.

1. Introduction

Classification identifies the class, from a set of classes, to which a new observation belongs, based on the training data containing observations whose class labels are known. Due to its importance in many applications, statistical approaches have been extensively developed. To name but a few, principal component analysis (PCA, Turk and Pentland, 1991), Fisher's linear discriminant analysis (LDA, Fisher, 1936; Rao, 1948), partial least square approaches (PLS, Barker and Rayens, 2003), etc. have all been explored for classification. The common essence of these approaches is to find optimal projections based on a particular criterion for subsequent classification. While the data dimension is moderate, these approaches or their variants often work nicely. With the advent of modern technology and devices for collecting data, the dimension of data can become very high and may be intrinsically infinite, such as functional data; this requires the aforementioned approaches to be adapted. Motivated by the Fisher's LDA, we propose *sensible* functional LDA (sFLDA) to search the optimal projections for subsequent classification.

LDA aims to find ideal linear projections and performs classification on the projected subspace. Ideal projections are those maximizing the projected distances between classes while keeping the projected distances among subjects in the same class minimized. Take a p-dimensional case for example; mathematically the ideal projection is the eigenvector $\boldsymbol{b}$ in

$$\Sigma_W^{-1}\Sigma_B\boldsymbol{b} = \lambda\boldsymbol{b}, \tag{1.1}$$

where Σ_W^{-1} denotes the inverse of the within-class covariance matrix Σ_W, and Σ_B is the between-class covariance matrix that characterizes the variation of class means. Under classical multivariate settings, Σ_W is invertible. Please refer to Mardia et al. (1980) for the details of LDA. Due to its simplicity, LDA has been widely employed in many applications.

Extending (1.1) directly to functional data is tricky due to the noninvertible covariance operator. Specifically, the inverse of the covariance operator is unbounded if the functional data is in $\mathcal{L}_2$, which is commonly assumed in the functional data

* Correspondence to: 128 Academia Rd, Sec 2, Nankang, Taipei 115, Taiwan.
E-mail address: cirenjiang@stat.sinica.edu.tw (C.-R. Jiang).

https://doi.org/10.1016/j.csda.2018.04.005
0167-9473/

analysis literature (e.g., Hall et al., 2006; Li and Hsing, 2010; Delaigle and Hall, 2012, etc.). To elucidate our idea, let us first introduce notations. Suppose the data consists of c classes. Let X_k be an $\mathcal{L}_2$ stochastic process, defined on a finite compact interval $\mathcal{T}$, in class k with mean function μ_k and a common covariance function Γ_W. Mercer's theorem implies that the covariance function can be further decomposed as $\Gamma_W(s,t) = \sum_{j=1}^{\infty}\lambda_j\phi_j(s)\phi_j(t)$, where $\lambda_1 > \lambda_2 > \cdots > 0$, ϕ_j is the corresponding eigenfunction of λ_j and $\sum_{j=1}^{\infty}\lambda_j < \infty$. Functional principal component analysis (FPCA) corresponds to a spectral decomposition of the covariance and leads to the well-known Karhunen–Loève decomposition of the random function,

$$X_k(t) = \mu_k(t) + \sum_{j=1}^{\infty} A_{k,j}\phi_j(t), \tag{1.2}$$

where $A_{k,j} = \langle X_k - \mu_k, \phi_j\rangle$ is the jth principal component score, $\mathrm{E}(A_{k,j}) = 0$, $\mathrm{var}(A_{k,j}) = \lambda_j$ and $t \in \mathcal{T}$. Here $\langle\cdot,\cdot\rangle$ stands for the inner product in $\mathcal{L}_2$, i.e., $\langle a, b\rangle = \int_{\mathcal{T}} a(t)b(t)dt$ for $a, b \in \mathcal{L}_2(\mathcal{T})$. Since we do not assume completeness on $\{\phi_j\}_{j=1}^{\infty}$, $\mu_k(t) = \sum_{j=1}^{\infty}\langle\mu_k,\phi_j\rangle\phi_j(t)$ is not guaranteed; note that a set of infinite number of basis functions does not imply completeness, e.g., Theorem 2.4.18 in Hsing and Eubank (2015). Further, we do not impose any parametric assumptions on X_k other than smoothness conditions on μ_k and on Γ_W, which are quite common in functional data analysis (e.g., Rice and Silverman, 1991; Chiou et al., 2003; Hall et al., 2006, etc.)

To handle the unbounded Γ_W^{-1}, basis-based approaches approximate the functional data with a set of finite basis functions and turn the functional problem into a multivariate one. For example, Hall et al. (2001), Glendinning and Herbert (2003), Müller (2005), Leng and Müller (2006), and Song et al. (2008) performed classification based on FPCA; Preda et al. (2007) classified functional data by means of PLS; Berlinet et al. (2008), Rincón and Ruiz-Medina (2012), and Chang et al. (2014) developed approaches based on wavelets. However, doing so might lose crucial information for subsequent classification due to an inappropriate set of basis functions. Take a binary case for example, let $\mu_1(t) = \sin(2\pi t)$, $\mu_2(t) = -\mu_1(t)$, and $\phi_j(t) = \sqrt{2}\cos(2j\pi t)$ for $j = 1, \ldots, \infty$ and $t \in [0, 1]$; neither μ_1 nor μ_2 can be represented with $\{\phi_j\}_{j=1}^{\infty}$. Therefore, FPCA based approaches might not be a good choice for such a case as the information of μ_1 is completely gone after projections. This argument is substantiated with simulated data in Section 6.

To search the projections of sFLDA efficiently as well as sensibly, we exploit the data structure when designing the procedure. Specifically, $\mathcal{S}_B$, the space spanned by $\{\mu_k\}_{k=1}^{c}$, possesses the information of how to maximize the distances between the projected class centers while $\mathcal{S}_W$, the space spanned by $\{\phi_j\}_{j=1}^{\infty}$, contains the information of how to reduce the within-class variations. With these in mind and to properly handle the noninvertibility issue of Γ_W, we propose a sensible procedure to find the projections in $\mathcal{S}_0$ and in $\mathcal{S}_1$ sequentially, where $\mathcal{S}_1$ (resp. $\mathcal{S}_0$) is the projection of $\mathcal{S}_B$ on $\mathcal{S}_W$ (resp. $\mathcal{S}_W^{\perp}$, the orthogonal complement of $\mathcal{S}_W$). In particular, the projections in $\mathcal{S}_0$ can completely discriminate functional observations in different classes (see Theorem 4.4 for details) while those in $\mathcal{S}_1$ are the optimal projections of functional LDA. Most existing approaches do not appear to appreciate that the optimal linear projections could be a set of the projections obtained in $\mathcal{S}_0$ and in $\mathcal{S}_1$; this may be because it suffices to consider projections in either $\mathcal{S}_1$ or $\mathcal{S}_0$ for binary classification problems. Accordingly, our procedure is more general.

Despite the difference in sampling schemes, functional data and longitudinal data come from similar sources. Therefore, it is practical to develop unified approaches for them (e.g., Müller, 2005; Hall et al., 2006; Jiang and Wang, 2010, etc.). James and Hastie (2001) employed natural cubic splines to tackle the problem of sparsity. Wu and Liu (2013) applied the FPCA approach proposed in Yao et al. (2005) to reconstruct sparsely observed longitudinal data and performed robust support vector machine (SVM) on the reconstructed curves. This strategy leads to the same predicament as other FPCA based approaches mentioned earlier. The major challenge in extending Fisher's LDA to longitudinal data is to perform classification on a new subject with longitudinal observations. The sparsity and irregularity of the observations make the projections difficult. We propose an imputation approach based on a conditional expectation technique (in Section 5) to resolve the sparsity issue without losing the subtle information about the mean functions.

There exist other functional classification approaches under different considerations. To name a few, Ferraty and Vieu (2003) and Galeano et al. (2015) investigated distance-based approaches, Hastie et al. (1995) and Araki et al. (2009) developed regularized approaches, Epifanio (2008) proposed an approach to classify functional shapes, Ferraty and Vieu (2006) suggested various of approaches based on k nearest neighbors (kNN) and kernel classification, Cuevas et al. (2007) introduced a depth-based approach, Li et al. (2012) recommended a method based on DD-plots, and Delaigle and Hall (2013) developed a functional classification framework when the observations were fragments of curves.

The rest of this paper proceeds as follows. In the next section, the motivation and the framework of sFLDA are introduced. The proposed estimators and their asymptotic properties are provided in Sections 3 and 4, respectively. We propose an imputation approach for longitudinal data while performing projections in Section 5. In Section 6, simulation studies under three data configurations are conducted. In Section 7, our approach, along with some competitors, is applied to two real data examples. Conclusions and discussions are given in the last section. Appendices include the assumptions made for the asymptotics, some details for Section 2.1 and the leave-one-curve-out cross-validation (CV) formulas of bandwidth selections. All the proofs are contained in the supplementary material.

Please cite this article in press as: Chen L.-H., Jiang C.-R., Sensible functional linear discriminant analysis. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.04.005.