Journal of Multivariate Analysis 143 (2016) 208–232

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Non-asymptotic adaptive prediction in functional linear models

Élodie Brunel, André Mas, Angelina Roche *

I3M, Université Montpellier II, Place Eugène Bataillon, CC051, 34000 Montpellier, France

ARTICLE INFO

Article history:
Received 15 January 2013
Available online 28 September 2015

AMS 2000 subject classifications:
62J05
62G08

Keywords:
Functional linear regression
Functional principal component analysis
Mean squared prediction error
Minimax rate
Penalized contrast estimator
Model selection on random bases

ABSTRACT

Functional linear regression has recently attracted considerable interest. Many works focus on asymptotic inference. In this paper we consider in a non asymptotic framework a simple estimation procedure based on functional Principal Regression. It revolves in the minimization of a least square contrast coupled with a classical projection on the space spanned by the m first empirical eigenvectors of the covariance operator of the functional sample. The novelty of our approach is to select automatically the crucial dimension m by minimization of a penalized least square contrast. Our method is based on model selection tools. Yet, since this kind of methods consists usually in projecting onto known non-random spaces, we need to adapt it to empirical eigenbasis made of data-dependent – hence random – vectors. The resulting estimator is fully adaptive and is shown to verify an oracle inequality for the risk associated to the prediction error and to attain optimal minimax rates of convergence over a certain class of ellipsoids. Our strategy of model selection is finally compared numerically with cross-validation.

Introduction

Functional data analysis has known recent advances in the past two decades, addressing simultaneously many fields of applications. We refer to Ferraty and Vieu [24] and Ramsay and Silverman [37] for detailed examples in medicine, linguistics and chemometrics and to Preda and Saporta [35] for applications in econometrics.

In this paper we suppose that the dependence between a real-valued response Y and a functional predictor X belonging to a Hilbert space $(\mathbf{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ is given by the functional linear model, namely

$$Y = \langle \beta, X\rangle + \varepsilon, \tag{1}$$

where ε stands for a noise term with variance σ^2 and is independent of X and $\beta \in \mathbf{H}$ is an unknown function to be estimated. In order to simplify the notations, the random variable X is supposed to be centred as well, which means that the function $t \mapsto \mathbf{E}[X(t)]$ is identically equal to zero.

By multiplying both sides of Eq. (1) by $X(s)$ and taking the expectation, we see easily that the function β is solution of

$$\Gamma\beta := \mathbf{E}\left[\langle \beta, X\rangle X(\cdot)\right] = \mathbf{E}\left[YX\right] =: g, \tag{2}$$

* Corresponding author.
E-mail addresses: ebrunel@math.univ-montp2.fr (É. Brunel), andre.mas@univ-montp2.fr (A. Mas), angelina.roche@univ-montp2.fr (A. Roche).

http://dx.doi.org/10.1016/j.jmva.2015.09.008
0047-259X/

where Γ is the covariance operator associated to the functional predictor X. Eq. (2) is known to be an ill-posed inverse problem (see Engl et al. [23, Chapter 2.1]).

The literature on the functional linear model is wide and numerous estimation procedures exist. A first method consists in minimizing a least square criterion subject to a roughness penalty. For instance, Li and Hsing [31] proposed an estimation procedure by minimization of such a criterion on periodic Sobolev spaces, Crambes et al. [19] generalized the well-known smoothing-spline estimator used in univariate nonparametric regression. Another approach is based on dimension reduction: this consists in approximating the regression function β by projection onto finite-dimensional spaces. Those spaces are usually obtained by taking the first components of a basis of $\mathbf{H}$. Some authors considered projection onto fixed basis, such as B-spline basis (Ramsay and Dalzell [36]) or general orthonormal basis (Cardot and Johannes [15]). But the most popular method is Functional Principal Component Regression (FPCR), this consists in taking the random space spanned by the eigenfunctions associated to the largest eigenvalues of the empirical covariance operator:

$$\Gamma_n : f \in \mathbf{H} \mapsto \frac{1}{n}\sum_{i=1}^{n}\langle f, X_i\rangle X_i. \tag{3}$$

The resulting estimator is shown to be consistent, but its behaviour is often erratic in simulation studies, thus a smooth version by using splines has been proposed by Cardot et al. [14]. The FPCR estimator is shown to attain optimal rates of convergence for the risk associated to the prediction error over fixed curves x (see Cai and Hall [11]) as well as for the $\mathbf{L}^2$-risk (see Hall and Horowitz [26]).

All the proposed estimators rely on the choice of at least one tuning parameter (the smoothing parameter appearing in the penalized criterion or the dimension of approximation space) which influences significantly the quality of estimation. Optimal choice of such parameters depends generally on both unknown regularities of the slope function β and the predictor X (see e.g. [11,19,15]) and the parameters are usually chosen in practice by cross-validation.

Until the recent work of Comte and Johannes [17], nonasymptotic results providing adaptive data-driven estimators were missing. Comte and Johannes [17,18] propose model selection procedures for the orthogonal series estimator introduced first by Cardot and Johannes [15]. In [17], they propose to select the dimension by minimization of a penalized contrast criterion under strong assumption of periodicity of the curve X while in [18] they define a dimension selection criterion by means of a stochastic penalized contrast emulating Lepski's method (see Goldenshluger and Lepski [25]) and do not require specific assumptions on the curve X. The resulting estimators are completely data-driven and achieve optimal minimax rates for general weighted $\mathbf{L}^2$-risks. However, since both dimension selection criteria depend on weights defining the risk, these selection procedures do not address prediction error, which can be written as a weighted norm whose weights are the unknown eigenvalues of the covariance operator.

In the same context as Comte and Johannes [17], Brunel and Roche [10] propose to estimate the slope function by minimizing a least square contrast on spaces spanned by the trigonometric basis. The dimension is selected by means of a penalized contrast. Their estimator is proved to attain the optimal minimax rate of convergence for the risk associated to the prediction error.

Another approach is proposed by Cai and Yuan [12] carrying out reproducing kernel Hilbert spaces. They develop a data-driven choice of the tuning parameter of the roughness regularization method (see e.g. Ramsay and Silverman [37]). Their estimation procedure is shown to attain the optimal rate of convergence without the need of knowing the covariance kernel. Lee and Park [30] also suggest general variable selection procedures based on a weighted L_1 penalty under assumption of sparsity on the functional parameter β. Their estimator is shown to be consistent and to satisfy the oracle-property.

In this paper, we propose an entirely data-driven procedure to select the adequate dimension for the classical FPCR estimator. The method proposed is based on model selection tools developed in a general context by Barron et al. [4], outlined by Massart [34], and in a context of regression by Baraud [2,3]. However, these tools are not meant to deal with estimators defined on random approximation spaces and thus have to be adapted. Section 1 is devoted to the description of estimation procedure. The resulting estimator is proved to satisfy an oracle-type inequality and to attain the optimal minimax rate of convergence for the risk associated to the prediction error for slope functions belonging to Sobolev classes in Section 2. In Section 3, a simulation study is presented including a comparison with cross-validation. The proofs are detailed in Appendix A.

1. Definition of the estimator

We assume that we are given an i.i.d. sample $(Y_i, X_i)_{i\geq 1}$ where the generic Y is real and X belongs to the Hilbert space $\mathbf{H}$. Thereafter, the Hilbert space is set to be $\mathbf{H} = \mathbf{L}^2([0,1])$ equipped with its usual inner product $\langle\cdot,\cdot\rangle$ defined by $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$ but our method adapts to more general Sobolev spaces as well. We assumed above that X is a centred random curve.

We recall that the theoretical covariance operator Γ of X defined by Eq. (2) in the introductory section is a selfadjoint trace class operator defined on and with values in $\mathbf{L}^2([0,1])$. This means that the sequence of its eigenvalues denoted $(\lambda_j)_{j\geq 1}$ is positive and summable. The associated sequence of eigenfunctions is denoted by $(\psi_j)_{j\geq 1}$.