Econometrics and Statistics 1 (2017) 150–166

Contents lists available at ScienceDirect

Econometrics and Statistics

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

Multinomial functional regression with wavelets and LASSO penalization

Seyed Nourollah Mousavi[a], Helle Sørensen[b,*]

[a] Department of Mathematics, Faculty of Sciences, Arak University, Arak 38156-8-8349, Iran
[b] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen East, Denmark

A R T I C L E I N F O

Article history:
Received 18 March 2016
Revised 26 August 2016
Accepted 25 September 2016
Available online 9 November 2016

Keywords:
Discrete wavelet transform
Functional predictor
Supervised classification
Lameness data for horses
Phoneme data

A B S T R A C T

A classification problem with a functional predictor is studied, and it is suggested to use a multinomial functional regression (MFR) model for the analysis. The discrete wavelet transform and LASSO penalization are combined for estimation, and the fitted model is used for classification of new curves with unknown class membership. The MFR approach is applied to two datasets, one regarding lameness detection for horses and another regarding speech recognition. In the applications, as well as in a simulation study, the performance of the MFR approach is compared to that of other methods for supervised classification of functional data, and MFR performs as well or better than the other methods.

1. Introduction

This paper is about classification for functional data. We consider situations with functional predictors where the aim is to classify new functions into well-specified groups. We propose a method based on multinomial functional regression (MFR) which, apart from the classification itself, also gives us information about which parts of the signals are used in the classification procedure. The multinomial regression approach is a generalization of functional logistic regression to multiclass problems, and estimation of the model combines wavelet expansions with LASSO regularization.

Our main application is concerned with diagnosis of lameness for horses. This is a difficult task even for experienced veterinarians, and we examine if acceleration signals collected during trot can be used as predictor. Data are available from eight horses, each tested in healthy condition and with lameness on either of the four limbs, and the aim is to classify new acceleration signals into lameness groups (healthy or not, and location of injury). Another application comes from speech recognition where the aim is to predict which phoneme is spoken, based on a log-periodogram (Hastie et al., 1995).

There are several approaches in the literature to classification of functional data. Early work include Hall et al. (2001) and James and Hastie (2001) who used linear discriminant analysis (LDA) on scores from a principal component analysis (PCA) and on coefficients from spline expansions, respectively, and Ferraty and Vieu (2003) using a kernel approach. Later, PCA was combined with logistic regression for the case with two groups (Müller and Stadtmüller, 2005), the method of partial least squares (PLS) was accommodated to functional data (Preda et al., 2007), and methods based on functional depth were suggested (Cuevas et al., 2007; López-Pintado and Romo, 2006). Recently Tian and James (2013) suggested a dimension

* Corresponding author.
E-mail addresses: n-mousavi@araku.ac.ir (S.N. Mousavi), helle@math.ku.dk (H. Sørensen).

http://dx.doi.org/10.1016/j.ecosta.2016.09.005
2452-3062/

reduction approach that takes into account the association to the categorical variable, and Delaigle and Hall (2012) studied optimality properties of a nearest centroid classifier.

Another topic in functional data analysis is regression with functional outcome and/or predictors. The situation with scalar response and functional covariates is of particular interest for this paper. In the simplest case we observe for each subject i a one-dimensional continuous response Y_i and a function $X_i : (0, 1) \to \mathbb{R}$, and assume (among others) that the conditional expectation of Y_i given $X_i = x_i$ is given by

$$\mathrm{E}[Y_i|X_i = x_i] = \alpha + \int_0^1 \beta(t)x_i(t)\,dt, \tag{1}$$

where α is an unknown intercept and $\beta : (0, 1) \to \mathbb{R}$ is an unknown coefficient function.

Several estimation approaches have been suggested for this model. One method, often referred to as functional principal component regression (FPCR), consists of a functional principal component analysis of the x_i's followed by a regression on the first few, say K, scores (Cardot et al., 1999; Ramsay and Silverman, 2005). This yields coefficient functions in the space spanned by the first K principal components (PCs), and it is thus implicitly assumed that these PCs not only account for a large proportion of the variation between X_i's, but are also relevant for the association between Y and X. Lee and Park (2012) discussed a selection approach to choose the most informative PC basis using LASSO, i.e., imposing a L^1 penalty on β, and the effect of a quadratic penalty on β in FPCR was discussed by Randolph et al. (2012).

Another approach is to use a rich, flexible basis for β in combination with regularization methods. For example, Marx and Eilers (1999) and Cardot et al. (2003) used spline series expansions and added penalty terms to the log-likelihood function, and Goldsmith et al. (2011) and Wood (2011) used spline series expansions in a mixed-model set-up. Reiss and Ogden (2007) combined FPCR, functional partial least squares, and penalized splines. Zhao et al. (2012) combined wavelet expansions with LASSO regression and is of particular importance for this paper. The combination is efficient since LASSO penalization by construction selects sparse models, and wavelets are known to offer sparse, yet precise, representations of many types of functions. LASSO has also been used in combination with other basis systems in order to obtain sparse representations (James et al., 2009; Lee and Park, 2012).

Many of the above-mentioned methods also apply to exponential families, in particular to the case with binary response leading to functional logistic regression (Cardot and Sarda, 2005; Crainiceanu et al., 2009; Goldsmith et al., 2011; James, 2002; Müller and Stadtmüller, 2005). Most of these papers contain asymptotic results but there are only few examinations of finite-sample properties in non-Gaussian cases. An exception is the paper by Reiss et al. (2015) where Gaussian and logistic regression with image predictors are studied.

We will take the logistic regression set-up a step further and consider multinomial regression with functional covariates. Let X_i be as before, but consider categorical outcomes Y_i with M possible outcomes, $m \in \mathcal{M}$. Define $p_m(x)$ as the conditional probability of class m given the functional outcome,

$$p_m(x) = P(Y = m|X = x), \quad m \in \mathcal{M},$$

and assume that $p_m(x)$ is proportional to $\exp(\alpha_m + \int_0^1 \beta_m(t)x(t)\,dt)$ for class-specific intercepts α_m and class-specific coefficient functions β_m. Once the model has been fitted, it can be used for classification in the obvious way: Given a curve x, compute $\hat{p}_m(x)$ for all m and allocate the curve to the group with highest probability.

We will follow the approach from Zhao et al. (2012) closely regarding estimation. More specifically, we select a family of wavelet bases and a resolution level, expand the covariate functions in the basis and use the wavelet coefficients as covariates in a multinomial regression with LASSO penalization. The LASSO tuning parameter and resolution level are selected by cross validation. The regression coefficients from the optimal multinomial regression are extracted and translated into estimated coefficient functions, $\hat{\beta}_m$.

Our main contribution versus Zhao et al. (2012) is the generalization from continuous to multinomial response, and although the model itself is straight-forward, the change of outcome type gives new challenges. First, the least angle regression (LARS) algorithm for the penalized regression problem used by Zhao et al. (2012) does not seem to be implemented for the multinomial case; instead we use a method based on coordinate descent (Friedman et al., 2010). Second, methods for supervised classification must be evaluated differently than prediction methods for numerical outcomes, and we examine the success rate for lameness diagnosis and speech recognition when MFR is used for classification. Finally, we investigate the specific choice of both vanishing moments and detail level for the wavelet basis more thoroughly than Zhao et al. (2012).

The rest of the paper is organized as follows. We go through the details about MFR in Section 2. The data on lameness are described and analyzed in Section 3, and Section 4 presents a simulation study inspired by the lameness data. In Section 5 the phoneme data are classified using MFR. Finally, we discuss the results and conclude in Section 6.

2. Multinomial functional regression

This section gives details about the model and the estimation procedure. Data consist of pairs (x_i, y_i) which are assumed to be outcomes from independent random variables (X_i, Y_i), $i = 1, \ldots, n$. The response variable Y_i is nominal with more than two levels indexed by $m \in \mathcal{M}$, and the explanatory variable X_i is a real-valued function defined on the unit interval (for simplicity), i.e., $X_i : (0, 1) \to \mathbb{R}$.