SUPERVISED CLASSIFICATION OF CURVES BY FUNCTIONAL DATA ANALYSIS: AN APPLICATION TO NEUROMARKETING DATA

Salvatore Latora [1] and Luigi Augugliaro[1]

[1] Department of Economics, Business and Statistics, University of Palermo, (e-mail: salvatore.latora@unipa.it, luigi.augugliaro@unipa.it)

ABSTRACT: In this paper we contribute to the functional data analysis literature by presenting a scalar-on-function penalized regression model with a multinomial response variable which takes into account possible information given by the phase variability. We also providing a practical application on neuromarketing data.

KEYWORDS: functional data, high-dimensional data, machine learning, sparse inference, supervised learning classification

1 Introduction

In recent decades, functional data analysis has played an increasingly important role in various scientific field, such as medicine, biology, engineering, and, above all, in the field of statistical research (see Ramsay & Silverman, 1997, Hsing & Eubank, 2015, Koner & Staicu, 2023 for some reference review). In this paper, we consider an application to neuromarketing data. *Neuromarketing* (Fisher *et al.*, 2010) is the application of neuroscientific methods to understand and analyse human behaviour in relation to markets and business needs. On the basis of different neurometrics, obtained by EEG recordings, taken on a sample of subjects while watching positive, negative, and neutral valence videos, to measure the α-*asymmetry* of the brain (a condition indicating the subject's attention to what he or she is observing, see Mazza & Pagano, 2017), the proposed methodology in this article aims to classify the valence of the video observed. The remaining part od this paper is organized as follows: in Section 2 we explain our proposal; in Section 3, the results obtained by analyzing the data introduced above are illustrated; Finally, conclusions are provided in Section 4.

2 Proposed model

Notation and definitions. By functional data, we mean a realization of a stochastic process. The functional data, i.e. the predictor, is modelled as: $f_{it_k} = f_i(t_k) + \varepsilon_{it_k}$, with $f_i \in \mathcal{F}$, where t_{ik} is the k-th time point detected on the i-th subject, with domain $[0,1]$, ε_{it_k} is an error term normally distributed, and f_{it_k} is an element of $\mathbb{L}^2_{[0,1]}$, where $\mathbb{L}^2_{[0,1]}$ denotes the space of square-integrable functions endowed with the standard inner product $\langle g_1\ ,\ g_2\rangle = \int_0^1 g_1(t)\, g_2(t)\, dt$ and associated norm $\|g\| = \langle g\ ,\ g\rangle^{\frac{1}{2}}$. Let us denote by Y_i, for $i = 1,\ldots,n$, a random variable distributed according to a Multinomial distribution, such that $Y_i \in \{-1,0,1\}$. Finally, by γ we denote a diffeomorphism, (*warping function*), belonging to the set $\Gamma = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0, \gamma(1) = 1\}$.

The propose model. The multinomial scalar-on-function regression model, belonging to the class of *FGLM* (James, 2002), takes the following form

$$\log\left\{\frac{Pr(Y_i = g \mid f_{it_k})}{Pr(Y_i = 0 \mid f_{it_k})}\right\} = \eta_{ig} = \beta_{0g} + \langle f_i\ ,\ \beta_g\rangle, \tag{1}$$

where β_{0g} is the intercept of the g-th group and $\beta_g \in \mathbb{L}^2(t)$ is the regression coefficient function. Usually, for classification purposes, the phase variability of functional data is not taken into account, making it unitary during the pre-processing step through time warping (Ramsay & Silverman, 1997). However, as some authors show, (e.g., see Tucker *et al.*, 2013) phase variability may contain useful information for classification purposes. In this setting, time is expressed as $t_{ik} = \gamma_i(t_k)$, where $\gamma_i \in \Gamma$ is the warping function. Hence, the functional predictor to be used in (1) is expressed in a new re-parametrization of time as $f_{it_k} = f_i(\gamma_i(t_k)) = \tilde{f}_i(t_k)$, where $\tilde{f}_i(t_k) \in \mathbb{L}^2_{[0,1]}$ which only contains information on amplitude variability. Therefore, to use both phase and amplitude variability for our prediction problem, model (1) becomes

$$\log\left\{\frac{Pr(Y_i = g \mid f_{it_k})}{Pr(Y_i = 0 \mid f_{it_k})}\right\} = \beta_{0g} + \langle \tilde{f}_i\ ,\ \beta_g\rangle + \langle \gamma_i\ ,\ \theta_g\rangle, \tag{2}$$

where $\langle \gamma_i\ ,\ \theta_g\rangle$ is the term contain information on the phase variability. Assuming that, both $\tilde{f}_i$ and γ_i are zero mean functions, and using by Karhunen–Loève expansion (Hsing & Eubank, 2015), i.e., $\tilde{f}_i(t) = \sum_{j=1}^{+\infty} X_{ij}\phi_j^f(t)$, and
$\gamma_i(t) = \sum_{l=1}^{+\infty} Z_{il}\phi_l^\gamma(t)$. Model (2) can be expressed as follow:

$$\eta_{ig} = \beta_{0g} + \sum_{j=1}^{p} X_{ij}\langle \phi_j^f\ ,\ \beta_g\rangle + \sum_{l=1}^{q} Z_{il}\langle \phi_l^\gamma\ ,\ \theta_g\rangle, \tag{3}$$

where X_{ij} and Z_{il} are the *scores*, obtained by *FPCA*. In our application we use the *PACE* method (Yao *et al.*, 2005). The model becomes a classic multinomial regression model on scores, in which there are high dimensionality problems due to the choice of the number of basis by which to approximate both $\tilde{f}_i$ and γ_i. To overcome the problems from the high dimensional setting, we propose to minimize the penalised log-likelihood function $l_\lambda(\mathbf{b}) = l(\mathbf{b}) + n\lambda P(\mathbf{b})$, where $\mathbf{b}$ denote a vector of parameters for both amplitude and phase variability terms, whereas λ is the tuning parameter and $P(\mathbf{b})$ is the *Elastic-Net* penalty function (Zou & Hastie, 2005), i.e.: $P(\mathbf{b}) = \alpha\|\mathbf{b}\|_1 + \frac{(1-\alpha)}{2}\|\mathbf{b}\|_2^2$.

3 Application to Neuromarketing Data

The sample consists of $n = 60$ subjects who participated to a study, in which each subject was shown a video having positive, neutral, or negative valence. Through EEG signals, two indices, BIS and BAS (Davidson *et al.*, 1990), were obtained capable of capturing whether the subject showed attention when viewing the video. In the preprocessing step, all the curves were aligned. Subsequently, four separate FPCAs for each indicator and related warping functions were made to obtain the scores.

Table 1. *Hyper parameter values and model performance metrics on test set.*

α	λ	Accuracy	Precision[a]	Recall[a]
0.9797	0.0045	0.933	0.944	0.933

[a] Macro average was used

Table 1 shows the selected hyper-parameter: the selected α parameter allowed for a very selective model, which leads to a Lasso-type penalty function, however, the selected λ value is close to zero. Again Table 1 shows how the model achieves almost perfect classification ability on the test set, and thus excellent generalization ability.

4 Conclusions

The proposed approach allows the extraction and selection of relevant signals for classification, also taking into account the possible information of phase variability through a specific term in the linear predictor. The results show in

Section 3 highlight that the proposed model has achieved an excellent degree of generalization.

References

DAVIDSON, R. J., EKMAN, P., SARON, C. D., SENULIS, J. A., & FRIESEN, W. V. 1990. Approach-Withdrawal and Cerebral Asymmetry: Emotional Expression and Brain Physiology. I. *Journal of Personality and Social Psychology*, **58**(2), 330–341.

FISHER, CARL ERIK, CHIN, LISA, & KLITZMAN, ROBERT. 2010. Defining Neuromarketing: Practices and Professional Challenges. *Harvard Review of Psychiatry*, **18**(4), 230–237.

HSING, TAILEN, & EUBANK, RANDALL. 2015. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. First edn. Wiley Series in Probability and Statistics. Wiley.

JAMES, GARETH M. 2002. Generalized Linear Models with Functional Predictors. *Journal of the Royal Statistical Society: Series B*, **64**(3), 411–432.

KONER, SALIL, & STAICU, ANA-MARIA. 2023. Second-Generation Functional Data. *Annual Review of Statistics and Its Application*, **10**(1), 547–572.

MAZZA, VERONICA, & PAGANO, SILVIA. 2017. *Electroencephalographic Asymmetries in Human Cognition*. New York, NY: Springer New York. Pages 407–439.

RAMSAY, J. O., & SILVERMAN, B. W. 1997. *Functional Data Analysis*. Springer Series in Statistics. New York, NY: Springer New York.

TUCKER, J. DEREK, WU, WEI, & SRIVASTAVA, ANUJ. 2013. Generative Models for Functional Data Using Phase and Amplitude Separation. *Computational Statistics & Data Analysis*, **61**(May), 50–66.

YAO, FANG, MÜLLER, HANS-GEORG, & WANG, JANE-LING. 2005. Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, **100**(470), 577–590.

ZOU, HUI, & HASTIE, TREVOR. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, **67**(2), 301–320.