Statistica Sinica **33** (2023), 1-30
doi:https://doi.org/10.5705/ss.202020.0214

COPULA-BASED FUNCTIONAL BAYES CLASSIFICATION WITH PRINCIPAL COMPONENTS AND PARTIAL LEAST SQUARES

Wentian Huang and David Ruppert

Cornell University

Abstract: We present a new functional Bayes classifier that uses principal component (PC) or partial least squares (PLS) scores from the common (i.e. pooled) covariance function, that is, the covariance function marginalized over groups. When the groups have different covariance functions, the PC or PLS scores need not be independent or even uncorrelated. We use copulas to model the dependence. Our method is semiparametric; the marginal densities are estimated nonparametrically using kernel smoothing, and the copula is modeled parametrically. We focus on Gaussian and t-copulas, but other copulas can be used. The strong performance of our methodology is demonstrated through simulation, real-data examples, and asymptotic properties.

Key words and phrases: Asymptotic theory, Bayes classifier, functional data, perfect classification, rank correlation, semiparametric model.

1. Introduction

Functional classification, where the features are continuous functions on a compact interval, is receiving increasing interest in fields such as chemometrics, medicine, economics, and environmental science. James and Hastie (2001) extended the linear discriminant analysis (LDA) to functional data (FLDA), including the case where the curves are partially observed. James (2002) proposed a functional version of the generalized linear model (FGLM), including functional logistic regression. Thereafter, the FGLM was further researched by, among others, Müller and Stadtmüller (2005), Li, Wang and Carroll (2010), Zhu, Vannucci and Cox (2010), McLean et al. (2014), and Shang and Cheng (2015). Aside from the FGLM, other classifiers have also been studied. Rossi and Villa (2006) applied support vector machines (SVM) to classify infinite-dimensional data. Cuevas, Febrero and Fraiman (2007) explored the classification of functional data based on data depth. Li and Yu (2008) suggested a functional segmented discriminant

Corresponding author: Wentian Huang, Department of Statistics and Data Science, Cornell University, Ithaca, NY 14850, USA. E-mail: wh365@cornell.edu.