Statistica Sinica **34** (2024), 1-20
doi:https://doi.org/10.5705/ss.202022.0057

OPTIMAL CLASSIFICATION FOR FUNCTIONAL DATA

Shuoyang Wang, Zuofeng Shang, Guanqun Cao* and Jun S. Liu

University of Louisville, New Jersey Institute of Technology, Michigan State University and Harvard University

Abstract: A central topic in functional data analysis is how to design an optimal decision rule, based on training samples, to classify a data function. We exploit the optimal classification problem in which the data functions are Gaussian processes. We derive sharp convergence rates for the minimax excess misclassification risk both when the data functions are fully observed and when they are discretely observed. We explore two easily implementable classifiers, based on a discriminant analysis and on a deep neural network, respectively, which both achieve optimality in Gaussian settings. Our deep neural network classifier is new in the literature, and demonstrates outstanding performance, even when the data functions are nonGaussian. For discretely observed data, we discover a novel critical sampling frequency that governs the sharp convergence rates. The proposed classifiers perform favorably in finite-sample applications, shown in comparisons with other functional classifiers in simulations and one real-data application.

Key words and phrases: Functional classification, functional deep neural network, functional quadratic discriminant analysis, Gaussian process, minimax excess misclassification risk.

1. Introduction

In many applications, data are collected in the form of functions, such as curves or images. Such data are referred to as functional data. A fundamental problem in functional data analysis is to classify a data function based on training samples. For instance, in the speech recognition data extracted from the TIMIT database (Ferraty and Vieu (2003)), the training samples are digitized speech curves of American English speakers from different phoneme groups, and the task is to predict the phoneme of a new speech curve. Classic multivariate analysis techniques, such as logistic regression or discriminant analysis, are not directly applicable, because functional data are intrinsically infinite-dimensional (Wang, Chiou and Müller (2016)). A common strategy is to adapt a multivariate analysis to functional settings, such as functional logistic regression (Araki et al. (2009)) and functional discriminant analysis (Shin (2008); Delaigle, Hall and Bathia (2012); Delaigle and Hall (2012, 2013); Galeano, Joseph and Lillo (2015); Dai, Müller and Yao (2017); Berrendero, Cuevas and Torrecilla (2018);

*Corresponding author.