Statistica Sinica **34** (2023), 1-19
doi:https://doi.org/10.5705/ss.202022.0258

FAST CONVERGENCE ON PERFECT CLASSIFICATION FOR FUNCTIONAL DATA

Tomoya Wakayama[1] and Masaaki Imaizumi*[1,2]

[1] *The University of Tokyo and* [2] *RIKEN AIP*

Abstract: We investigate perfect classification on functional data using finite samples. Perfect classification for functional data is easier to achieve than for finite-dimensional data, because a sufficient condition for the existence of a perfect classifier, called the Delaigle-Hall condition, is available only for functional data. However, a large sample size is required to achieve perfect classification, even when the Delaigle-Hall condition holds, because the minimax convergence rate of the errors with functional data has a logarithm order in the sample size. We resolve this complication by proving that the Delaigle-Hall condition also achieves fast convergence of the misclassification error in a sample size under the bounded entropy condition on functional data. We study a reproducing kernel Hilbert space-based classifier under the Delaigle-Hall condition, and show that the convergence rate of its misclassification error has an exponential order in the sample size. Technically, our proof is based on (i) connecting the Delaigle-Hall condition and a margin of classifiers, and (ii) handling metric entropy of functional data. The results of our experiments support our findings, and show that other classifiers for functional data have a similar property.

Key words and phrases: Convergence rate, functional data, perfect classification, reproducing kernel Hilbert space.

1. Introduction

The classification problem is one of the most general and significant problems in functional data analysis. The goal of classification is to predict labels or categories from functional data given in the form of (possibly) infinite-dimensional random curves. Because of its versatility, classification has many applications in the field of science (Varughese et al. (2015)), engineering (Gannaz (2014); Li et al. (2013); Florindo, de Castro and Bruno (2011)), medicine (Chang, Chen and Ogden (2014); Islam (2020); Dai, Müller and Yao (2017)), and others. Several methods have been developed to solve this problem, including distance-based (Alonso, Casado and Romo (2012); Ferraty and Vieu (2003)), k-nearest neighbor (Biau, Bunea and Wegkamp (2005); Cérou and Guyader (2006)), partially least square (Preda, Saporta and Lévéder (2007); Preda and Saporta (2005)), orthonormal basis (Delaigle and Hall (2013, 2012)), Bayesian (Wang and Qu

*Corresponding author.