Statistica Sinica **33** (2023), 1-22
doi:https://doi.org/10.5705/ss.202021.0018

AN ONLINE PROJECTION ESTIMATOR FOR NONPARAMETRIC REGRESSION IN REPRODUCING KERNEL HILBERT SPACES

Tianyu Zhang and Noah Simon

University of Washington

Abstract: The goal of nonparametric regression is to recover an underlying regression function from noisy observations, under the assumption that the regression function belongs to a prespecified infinite-dimensional function space. In the online setting, in which the observations come in a stream, it is generally computationally infeasible to refit the whole model repeatedly. As yet, there are no methods that are both computationally efficient and statistically rate optimal. In this paper, we propose an estimator for online nonparametric regression. Notably, our estimator is an empirical risk minimizer in a deterministic linear space, which is quite different from existing methods that use random features and a functional stochastic gradient. Our theoretical analysis shows that this estimator obtains a rate-optimal generalization error when the regression function is known to live in a reproducing kernel Hilbert space. We also show, theoretically and empirically, that the computational cost of our estimator is much lower than that of other rate-optimal estimators proposed for this online setting.

Key words and phrases: Mercer expansion, nonparametric regression, online learning, reproducing kernel Hilbert space.

1. Introduction

It is often of interest to estimate an underlying regression function, linking features to an outcome, from noisy observations. When the structure of this function is not known (e.g., when we do not want to assume a simple linear form), some form of nonparametric regression is employed. More formally, suppose we observe $(X_i, Y_i) \overset{i.i.d.}{\sim} \rho(X, Y)$, for $i = 1, 2, \ldots, n$, generated from the following statistical model:

$$Y_i = f_\rho(X_i) + \epsilon_i, \tag{1.1}$$

where, for each i, $X_i \overset{i.i.d.}{\sim} \rho_X$ (which take values in $\mathbb{R}^d$) are our features, $Y_i \in \mathbb{R}$ is our outcome, ϵ_i are independent and identically distributed (i.i.d.) mean zero

Corresponding author: Tianyu Zhang, Department of Biostatistics, University of Washington, Seattle, WA 98195, USA. E-mail: zty@uw.edu.