DOCTORAL THESIS

Statistical Learning for Kernel-Based Functional Linear Regression

GUO, Keli

Date of Award:
2022

Abstract

Over the last two decades, functional linear regression that relates a scalar response on a functional predictor has been extensively studied. In practice, however, apart from functional predictors, scalar predictors or outliers are frequently included in the dataset. To address this issue, we investigate three variants of the functional linear regression model within the framework of reproducing kernel Hilbert space (RKHS), respectively.

First, we consider the semi-functional linear model that consists of a functional component and a nonparametric component. A double-penalized least squares method is adopted to estimate both the functional and nonparametric components within the framework of reproducing kernel Hilbert space. By virtue of the representer theorem, an efficient algorithm that requires no iterations is proposed to solve the corresponding optimization problem, where the regularization parameters are selected by the generalized cross-validation criterion. Moreover, we establish minimax rates of convergence for prediction in the semi-functional linear model. Our results reveal that the functional component can be learned with the minimax optimal rate as if the nonparametric component were known. Numerical studies and real data analysis are provided to demonstrate the effectiveness of the method and to verify the theoretical findings.

Then we consider the partially functional linear regression model (PFLM) that consists of a functional linear regression component and a sparse high-dimensional linear regression component. We adopt a double-penalized least squares approach to estimate the functional component within the framework of reproducing kernel Hilbert space and the parametric component by sorted ℓ_1 penalized estimation (SLOPE). Moreover, we establish minimax rates of convergence for prediction in the PFLM. Our results suggest that the estimator obtained by SLOPE can achive the minimax optimal rate regardless of the functional component. In contrast, the learning rate for the functional component depends on both functional and parametric components. To solve the optimization problem, an efficient computing algorithm is proposed with the help of the representer theorem. Numerical studies are conducted to demonstrate the performance of the proposed method.

Finally, we propose an outlier-resistant functional linear regression model so that we can perform robust regression and outlier detection simultaneously. The proposed model includes a subject-specific mean shift parameter in the functional linear regression model to indicate whether an observation is an outlier or not. We adopt a double-penalized least squares method to estimate the functional component within the framework of reproducing kernel Hilbert space and the mean shift parameter by ℓ_1 penalization or SLOPE. By virtue of the representer theorem, an efficient algorithm is proposed to solve the corresponding optimization problem. Moreover, we establish the minimax rates of convergence for prediction and estimation in the proposed model. Our results reveal that the convergence rate for estimation of the mean shift parameter is not affected by the functional component. The functional component can be learned with the minimax optimal rate as if there were no outliers. Numerical studies are provided to demonstrate the effectiveness of the proposed methods.