BIOSTATISTICS SEMINAR

Optimal Nonparametric Inference with Two-Scale Distributional Nearest Neighbors

Jinchi Lv, PhD
Kenneth King Stonier Chair in Business Administration
Professor of Data Sciences and Operations, and Mathematics
University of Southern California

The weighted nearest neighbors (WNN) estimator has been popularly used as a flexible and easy-to-implement nonparametric tool for mean regression estimation. The bagging technique is an elegant way to form WNN estimators with weights automatically generated to the nearest neighbors; we name the resulting estimator as the distributional nearest neighbors (DNN) for easy reference. Yet, there is a lack of distributional results for such estimator, limiting its application to statistical inference. Moreover, when the mean regression function has higher-order smoothness, DNN does not achieve the optimal nonparametric convergence rate, mainly because of the bias issue. In this work, we provide an in-depth technical analysis of the DNN, based on which we suggest a bias reduction approach for the DNN estimator by linearly combining two DNN estimators with different subsampling scales, resulting in the novel two-scale DNN (TDNN) estimator. The two-scale DNN estimator has an equivalent representation of WNN with weights admitting explicit forms and some being negative. We prove that, thanks to the use of negative weights, the two-scale DNN estimator enjoys the optimal nonparametric rate of convergence in estimating the regression function under the fourth-order smoothness condition. We further go beyond estimation and establish that the DNN and two-scale DNN are both asymptotically normal as the subsampling scales and sample size diverge to infinity. For the practical implementation, we also provide variance estimators and a distribution estimator using the jackknife and bootstrap techniques for the two-scale DNN. These estimators can be exploited for constructing valid confidence intervals for nonparametric inference of the regression function. The theoretical results and appealing finite-sample performance of the suggested two-scale DNN method are illustrated with several simulation examples and a real data application. This is a joint work with Emre Demirkaya, Yingying Fan, Lan Gao, Patrick Vossler and Jingbo Wang.

Thursday, February 16, 2023, 3:30-4:30 PM Eastern

133 Rosenau Hall

Virtual using link and info below.

https://unc.zoom.us/j/91249030964?pwd=UXloTWlHajdQbkRqd1d5TnRaMitYdz09

Meeting ID: 912 4903 0964

Passcode: 884852