Nonparametric and Semiparametric Regression For Longitudinal/Clustered Data and High Dimensional Data

Xihong LIN, *Department of Biostatistics, University of Michigan, USA*, E-mail: xlin@umich.edu

Abstract: We consider nonparametric and semiparametric regression estimation for longitudinal/clustered data and high dimensional data. The first half of the talk focuses on nonparametric and semiparametric regression estimation for clustered/longitudinal data using kernel and spline (smoothing and regression) methods. We show that unlike independent data, kernels and splines are not asymptotically equivalent for clustered/longitudinal data. Conventional kernel extensions of GEEs fail to account for the within-cluster correlation, while spline methods are able to account for this correlation. However, we identify an asymptotic equivalent kernel for the smoothing spline. We show that the most efficient spline and the newly proposed seemingly related kernel estimators can effectively account for the within-cluster correlation. We extend the results to semiparametric regression models, where some covariate effects are modeled parametrically, while others are modeled nonparametricaly. We derive the semiparametric efficient score and show the profile/kernel or profile/spline estimator is semiparametric efficient. The second half of the talk considers semiparametric regression models for the small n big p problem, such as in microarray data, where a clinical outcome depends on clinical covariates parametrically and gene expressions (often large) nonparametrically. Estimation proceeds with the Support Vector Machine technique, a technique most familiar to the machine learning community. We show that there is a close connection between the support vector machine and a linear mixed effects model, and estimation can proceed within the linear mixed model framework. The results are illustrated using simulation studies and data examples.