

Machine Learning for Precision Medicine: Model Selection, Estimation, and Inference

Yi Li

University of Michigan, USA

Abstract - In the era of precision medicine, high-throughput data are routinely collected. These high dimensional data defy classical regression models, which are either infeasible to fit or likely to incur low predictability because of overfitting. In this talk we will introduce several cutting-edge machine learning methods, developed by my group in the last few years, for modeling (censored) outcome data with high dimensional predictors. Specifically, we will introduce a Dantzig selector for fitting survival models with high dimensional predictors, followed by various semiparametric and nonparametric feature screening methods for handling ultra-high dimensional predictors. We will also discuss statistical inference for regression models with high dimensional predictors. With high dimensional outcome data, we will introduce a new class of high dimensional Gaussian graphical regression models with predictors. The talk focuses on statistical principles and concepts behind these methods, which are motivated and illustrated by various biomedical examples, which have precision medicine contexts.