BIOSTATISTICS SEMINAR

Adaptive Spanning Tree/Forest-based Lasso: Learning Model Heterogeneity in the Spatial Network

Xin Zhang, PhD
Research Scientist at Meta (Facebook)

Model heterogeneity identification aims to simultaneously estimate model parameters and reveal the model dissimilarity/cluster pattern among observations. In recent years, this topic has attracted a surge of research interests and found important applications in many scientific fields, such as personalized medicine, image analysis, and economic study. In this talk, Dr. Zhang is going to introduce two novel fusion Lasso-based model heterogeneity identification approaches, which incorporate the spatial information and balance the trade-off of model estimation accuracy and computation complexity. Our first approach is the distributed spanning tree-based fused-Lasso regression (DTFLR) approach designed for data integration in the distributed network systems. The proposed DTFLR approach consists of two key components, adaptive spanning tree-based fusion penalty for the low-complexity model design and decentralized generalized ADMM for parallel computing. We show that the estimator in DTFLR satisfies elegant statistical oracle properties, and the algorithm enjoys a linear convergence rate. In the follow-up work, we further improve the approach to analyze spatial image data. To avoid the tremendous computation and maintain the estimation accuracy, we develop a fusion penalty termed as the forest Lasso. More specifically, the forest Lasso is an adaptive fusion penalty, of which the adaptive weights are from the average of the estimations over multiple spanning trees. We show that the proposed forest Lasso approach can significantly improve the accuracy while only demands a low computation cost. In the end, Xin Zhang will talk about some future extensions and other related research directions. This talk is based on the joint works with Prof. Zhengyuan Zhu (ISU), Prof. Jia Liu (OSU), Prof Xin Wang (Miami), and Prof. Shan Yu (UVA).

Thursday February 24, 2022, 3:30-4:30 PM Eastern

McGavran-Greenberg Hall - Room 1301

Virtual using link and info below.

Link: https://unc.zoom.us/j/98412143955?pwd=a1p6c3hvZ28wSnk3dVVXQWl0dEpzdz09

Meeting ID: 984 1214 3955 Passcode: 0375501630

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH Biostatistics