University of California, Irvine
Statistics Seminar

Scalable Sparse Cox's Regression for Large-Scale Survival Data via Broken Adaptive Ridge

Gang Li
Professor of Biostatistics and Biomathematics
University of California, Los Angeles

Thursday, Nov. 30, 2017
4 p.m., 6011 Bren Hall
(Bldg. #314 on campus map)

Advancing medical informatics tools and high-throughput biological experimentation are making large-scale data routinely accessible to researchers, administrators, and policy-makers. This ``data deluge'' poses new challenges and critical barriers for quantitative researchers as existing statistical methods and software grind to a halt when analyzing these large-scale datasets, and calls for a need of methods that can readily fit large-scale data. In this talk I will present a new sparse Cox regression method for high-dimensional massive sample size survival data. Our method is an L0-based iteratively reweighted L2-penalized Cox regression, which inherits some appealing properties of both L0 and L2 penalized Cox regression while overcoming their limitations. We establish that it has an oracle property for selection and estimation and a grouping property for highly correlated covariates. We develop an efficient implementation for high-dimensional massive sample size survival data, which exhibits up to a 20-fold speedup over a competing method in our numerical studies. We also adapt our method to high-dimensional small sample size data. The performance of our method is illustrated using simulations and real data examples.