The 37th New England Statistics Symposium

SHORT COURSES

An Introduction to the Statistical Foundations of Transfer Learning

Full day; Hybrid, May 20, 2024

Dr. Yang Feng

Ye Tian

Instructor: Dr. Yang Feng is a professor of Biostatistics at New York University. Ye Tian is a fifth-year PhD student in Statistics at Columbia University.

Outline: This course offers a comprehensive introduction to the statistical foundations underpinning the prevalent machine learning technique: transfer learning. We delve into how transfer learning effectively transfers knowledge from one task to another in an adaptive and robust fashion, thereby enhancing model performance across both supervised and unsupervised learning frameworks.

Large Scale Data Science

Full day; Hybrid, May 20, 2024

Dr. Marc G. Genton

Dr. Sameh Abdulah

Instructor: Dr. Marc G. Genton is a Distinguished Professor of Statistics at the Spatio-Temporal Statistics and Data Science (STSDS) research group, King Abdullah University of Science and Technology (KAUST), Saudi Arabia. Dr. Sameh Abdulah is a research scientist at the Extreme Computing Research Center (ECRC), King Abdullah University of Science and Technology, Saudi Arabia. Dr. Mary Lai Salvaña is an Assistant Professor in Statistics at the University of Connecticut (UConn). .

Outline: The course content will cover the basic concepts of large-scale spatial statistics on parallel systems through synthetic and real data examples using both exact and approximation methods. The course will also provide a comprehensive comparison between existing Geostatistics packages (fields and GeoR) with the cutting-edge HPC packages (ExaGeoStatR and MPCR) to show the main contribution and benefits of using HPC techniques on leading-edge parallel hardware architectures such as GPUs and supercomputers.

Dr. Mary Lai Salvaña

Statistical Network Analysis in R

Half day - AM Session; Hybrid, May 20, 2024

Dr. Eric Kolaczyk

Instructor: Dr. Eric Kolaczyk is a professor in the Department of Mathematics and Statistics, and inaugural director of the McGill Computational and Data Systems Initiative (CDSI).

Outline: A gentle introduction to the statistical analysis of network data, largely through the lens of the R package igraph. Topics to be covered include basic definitions and concepts in networks, manipulation and visualization of network data, and tools for describing network characteristics, as well as a brief look at select inferential topics such as node clustering (aka 'community detection') and network modeling.

Website: https://symposium.nestat.org/short-courses.html