Stat364: Scalable Statistical Inference for Big Data with Applications
Time: 10:30am-11:45am, Monday, Spring 2024
Location: Science Center 706
Lead Instructor: Xihong Lin

Description: Unprecedented advances in digital technology have produced a revolution that is transforming science and society. Big data have been rapidly generated in many disciplines, such as genomics, health, physical and social sciences. The value of big data lies in effective analysis using statistical and machine learning methods that are computationally scalable and efficient. This seminar course is offered in response to these needs. Discussions will be co-led by several faculty members in the Departments of Statistics and Biostatistics at Harvard with active participation by students, postdoctoral fellows and faculty, with the goal of stimulating and discussing several cutting-edge ongoing research areas on scalable statistical and machine learning methods for big data. Students and postdoctoral fellows will work with faculty on group class presentations. The class is expected to have a lot of active discussions of current ongoing methodological research of faculty participants.

The discussions include current research activities, challenges and open problems. Examples include feature selection, manifold learning, representation learning, transfer learning, distributed and federated learning, ensemble learning and inference, topic modeling, Large Language Models (LMMs), and generative AI. Applications of these methods in various areas will be discussed.

Class schedule and reading list: http://bit.ly/stats364_presentation

Lecture slides and references: see the files in the class Google drive

Responsibilities:

- Active class participation
- Read papers in the reading list
- Serve as a discussant and present discussions by learning details, presenting additional materials, running software and data analysis.

Signup Sheet

http://bit.ly/stat364_signup