भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

The Decision Sciences Area at IIM Bangalore welcomes you to a webinar, titled:

Some novel screening and variable selection methods in ultra-high dimensional linear regression

by
Dr. Somak Dutta

Date: 20th October, 2021
Time: 6:30 p.m. to 7:30 p.m.

Abstract:

During the last couple of decades, substantial research has been devoted to identifying the important covariates in an ultra-high dimensional linear regression where the number of covariates is in the lower exponential order of sample size. While the notion of variable screening focuses on identifying a smaller subset of covariates that includes the important ones with overwhelmingly large probability, the notion of variable selection indulges only on identifying the truly important ones. Typically, because variable selection is costly, a screening step is performed to reduce the number of potential covariates. In this talk, we propose three new novel methodologies. For variable screening we first introduce the concept of ridge partial correlation to generalize the concept of partial correlation to ultra-high dimension setup. We then develop a sequential Bayesian rule to incorporate prior information on the true model size and effect sizes. Finally we propose a scalable variable selection method that embeds variable screening in its algorithm, thus providing scalability and alleviating the need of a two-stage method. Our theoretical results guarantee that under some mild conditions, we have screening consistency and selection consistency under ultra-high dimensional setup even when the error distribution is misspecified. We illustrate our methods using a dataset with close to half-a-million covariates. This talk is based on several joint works with Dr. Vivekananda Roy and PhD students Dongjin Li and Run Wang