Mathematics, Statistics, and Computer Science **@ UIC**

Statistics and Data Science Seminar

Model-free Feature Screening and FDR Control with Knockoff Features

Yuan Ke (University of Georgia)

Abstract: We proposes a model-free and data-adaptive feature screening method for ultra-high dimensional data. The proposed method is based on the projection correlation which measures the dependence between two random vectors. This projection correlation based method does not require specifying a regression model, and applies to data in the presence of heavy tails and multivariate responses. It enjoys both sure screening and rank consistency properties under weak assumptions. A two-step approach, with the help of knockoff features, is advocated to specify the threshold for feature screening such that the false discovery rate (FDR) is controlled under a pre-specified level. The proposed two-step approach enjoys both sure screening and FDR control simultaneously if the pre-specified FDR level is greater or equal to $1/s$, where s is the number of active features. The superior empirical performance of the proposed method is illustrated by simulation examples and real data applications.

Wednesday, October 21 at 4:00 PM in Zoom