

Data Science Seminar  
Hosted by Department of Mathematical Sciences

- Date: Tuesday, Nov 03, 2020
- Time: 12:00pm - 1:30pm
- Room: Via Zoom
- Speaker: Shaofei Zhao (Binghamton University)
- Title: Distribution-free and nonparametric multivariate feature screening via measure transportation for high dimensional response and predictor variables.

**Abstract**

Feature screening is an effective approach in selecting influential features from the explosion of big data with unprecedented dimensionality and complexity. Based on the integration of multivariate-rank via measure transportation and distance correlation, we propose a novel sure independence screening approach (MrDc-SIS). The MrDc-SIS achieves multiple desirable properties such as being exactly distribution-free, completely nonparametric, scale-free, robust for outliers or heavy tails, and sensitive for hidden structures. It represents an important advancement for real-world ultrahigh dimensional data that are messy in wide varieties. We establish the asymptotic sure screening consistency property under a mild condition by lifting any assumption about the finite moments. Moreover, the MrDc-SIS focuses on “large  $p$ , large  $q$ , small  $n$ ” and can be used to screen not only predictor variables like the majority of approaches in feature screening literature do, but also can screen response variables. Simulation studies demonstrate that MrDc-SIS performs better than other relevant approaches under various settings. We explore a challenging scenario when the number of responses ( $q = 10,000$ ) and the number of predictors ( $p = 10,000 \times 3$ ) are both much larger than the number of observations ( $n = 200$ ). We also applied the MrDc-SIS approach to a multi-omics lung cancer data from The Cancer Genome Atlas (TCGA).

**Note:** This is an ABD Exam.

From:  
<https://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:  
<https://www2.math.binghamton.edu/p/seminars/datasci/201103>

Last update: **2020/10/29 14:54**

