

Statistics Seminar
Department of Mathematical Sciences

| | |
|------------------|---|
| DATE: | Thursday, October 12, 2017 |
| TIME: | 1:15pm - 2:15pm |
| LOCATION: | WH G02 (note special location) |
| SPEAKER: | Jianqing Fan, Princeton University |
| TITLE: | Distributed Estimation of Principal Eigenspaces |

This talk is part of the Dean's Speaker Series in Statistics and Data Science.

Abstract

Principal component analysis (PCA) is fundamental to statistical machine learning. It extracts latent principal factors that contribute to the most variation of the data. When data are stored across multiple machines, however, communication cost can prohibit the computation of PCA in a central location and distributed algorithms for PCA are thus needed. This paper proposes and studies a distributed PCA algorithm: each node machine computes the top eigenvectors and transmits them to the central server; the central server then aggregates the information from all the node machines and conducts a PCA based on the aggregated information. We investigate the bias and variance for the resulting distributed estimator of the top K eigenvectors. In particular, we show that for distributions with symmetric innovation, the distributed PCA is "unbiased". We derive the rate of convergence for distributed PCA estimators, which depends explicitly on the effective rank of covariance, eigen-gap, and the number of machines. We show that when the number of machines is not unreasonably large, the distributed PCA performs as well as the whole sample PCA, even without full access of whole data. The theoretical results are verified by an extensive simulation study. We also extend our analysis to the heterogeneous case where the population covariance matrices are different across local machines but share similar top eigen-structures.

(Joint with Dong Wang, Kaizheng Wang, Ziwei Zhu)

About the speaker: Jianqing Fan, is a statistician, financial econometrician, and data scientist. He is Frederick L. Moore '18 Professor of Finance, Professor of Statistics, and Professor of Operations Research and Financial Engineering at the Princeton University where he chaired the department from 2012 to 2015. He is the winner of The 2000 COPSS Presidents' Award, Morningside Gold Medal for Applied Mathematics (2007), Guggenheim Fellow (2009), Pao-Lu Hsu Prize (2013) and Guy Medal in Silver (2014). He got elected to Academician from Academia Sinica in 2012. Fan is interested in statistical theory and methods in data science, statistical machine learning, finance, economics, computational biology, biostatistics with particular skills on high-dimensional statistics, nonparametric modeling, longitudinal and functional data analysis, nonlinear time series, wavelets, among others.

From:

<http://www2.math.binghamton.edu/> - **Binghamton University Department of Mathematical Sciences**

Permanent link:

<http://www2.math.binghamton.edu/p/seminars/stat/171012>

Last update: **2017/09/28 13:06**

