

**The Hong Kong Polytechnic University
Department of Applied Mathematics**

AMA Distinguished Seminar Series in Data Science and Machine Learning

High-dimensional Clustering via A Latent Semiparametric Mixture Model

By

**Prof. Hui ZOU
University of Minnesota**

Abstract

Cluster analysis is a fundamental task in machine learning. Several clustering algorithms have been extended to handle high-dimensional data by incorporating a sparsity constraint in the estimation of a mixture of Gaussian models. Though it makes some neat theoretical analysis possible, this type of approach is arguably restrictive for many applications. In this work we propose a novel latent variable transformation mixture model for clustering in which we assume that after some unknown monotone transformations the data follows a mixture of Gaussians. Under the assumption that the optimal clustering admits a sparsity structure, we develop a new clustering algorithm named CESME for high-dimensional clustering. The use of unspecified transformation makes the model far more flexible than the classical mixture of Gaussians. On the other hand, the transformation also brings quite a few technical challenges to the model estimation as well as the theoretical analysis of CESME. We offer a comprehensive analysis of CESME including identifiability, initialization, algorithmic convergence, and statistical guarantees on clustering. Extensive numerical study and real data analysis show that CESME outperforms the existing high-dimensional clustering algorithms including CHIME, sparse spectral clustering, sparse K-means, sparse convex clustering, and IF-PCA.

Biography

Hui Zou received PhD in Statistics from Stanford University in 2005 and became a full professor of Statistics at University of Minnesota in 2014. His research interests include high-dimensional statistics, machine learning, and data science in general. Some of his well-known inventions include elastic net, adaptive lasso, sparse PCA, LLA for nonconvex regularization, composite quantile regression, rank-based inference of graphical model, and high-dimensional discriminant analysis. He is an elected Fellow of Institute of Mathematical Statistics and American Statistical Association. He has been Web of Science highly cited researcher in mathematics from 2014 to 2019. He received several other awards including NSF career award, IMS Tweedie award and the best paper award in 2019 ICCM.

Date: 17 May 2022 (Tuesday)

Time: 9:30-10:30 (Hong Kong Standard Time GMT +8)

Venue: Online Talk via Zoom (Meeting ID: 940 2951 4899)

Speaker: Prof. Hui Zou, University of Minnesota

Host: Prof. Xingqiu Zhao, The Hong Kong Polytechnic University

Click to join:

<https://polyu.zoom.us/j/94029514899?pwd=K2tiUWxSNjZ5ZFUwekw0cXJiQUUpnUT09>



[Click to join \(Zoom\)](#)

*** * * ALL ARE WELCOME * * ***

For enrolment, please send your name and email to wai-yan.moon@polyu.edu.hk on or before 16 May 2022