



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

SEMINAR ON STATISTICS

Minimax-Optimal Dimension-Reduced Clustering for High-Dimensional Nonspherical Mixtures

By

Prof. Yuqi GU
Columbia University

Abstract

In mixture models, nonspherical (anisotropic) noise within each cluster is widely present in real-world data. We study both the minimax rate and optimal statistical procedure for clustering under high-dimensional nonspherical mixture models. In high-dimensional settings, we first establish the information-theoretic limits for clustering under Gaussian mixtures. The minimax lower bound unveils an intriguing informational dimension-reduction phenomenon: there exists a substantial gap between the minimax rate and the oracle clustering risk, with the former determined solely by the projected centers and projected covariance matrices in a low-dimensional space. Motivated by the lower bound, we propose a novel computationally efficient clustering method: Covariance Projected Spectral Clustering (COPO). Its key step is to project the high-dimensional data onto the low-dimensional space spanned by the cluster centers and then use the projected covariance matrices in this space to enhance clustering. We establish tight algorithmic upper bounds for COPO, both for Gaussian noise with flexible covariance and general noise with local dependence. Our theory indicates the minimax-optimality of COPO in the Gaussian case and highlights its adaptivity to a broad spectrum of dependent noise. Extensive simulation studies under various noise structures and real data analysis demonstrate our method's superior performance.

Date : 13 May 2025 (Tuesday)

Time : 10:00a.m. – 11:00a.m.

Venue : Room 1409 (Lift 25/26)

All are Welcome!