THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

SEMINAR ON STATISTICS AND DATA SCIENCE

Clustering via Uncoupled Regression (CURE)

By

Dr. Kaizheng WANG
Princeton University and Columbia University

Abstract

In this talk, we first consider a canonical clustering problem where one receives unlabeled samples drawn from a balanced mixture of two elliptical distributions and aims for a classifier to estimate the labels. Many popular methods including PCA and k-means require individual components of the mixture to be somewhat spherical, and perform poorly when they are stretched. To overcome this issue, we propose a non-convex program seeking for an affine transform to turn the data into a one-dimensional point cloud concentrating around -1 and 1, after which clustering becomes easy. Our theoretical contributions are two-fold: (1) we show that the non-convex loss function exhibits desirable landscape properties as long as the sample size exceeds some constant multiple of the dimension, and (2) we leverage this to prove that an efficient first-order algorithm achieves near-optimal statistical precision even without good initialization. We also propose a general methodology for multi-class clustering tasks with flexible choices of feature transforms and loss objectives.

Date   **: 08 May 2020 (Friday)**
Time   **: 9:30am – 11:00am**
Zoom Meeting **: https://hkust.zoom.us/j/5616960008**

All are Welcome!