

*Department of Mathematics,
University of California San Diego*

Final Defense

Xiao Pu

UCSD

Topics in Clustering: Feature Selection and Semiparametric Modeling

Abstract:

Clustering objects into similar clusters is an important practical problem in a wide variety of fields, including statistics, physics, bioinformatics, artificial intelligence, and data mining. My thesis focuses on feature selection and semiparametric modeling in clustering. In this talk, I will present three of the projects I have done with my advisor during my PhD studies. The first one proposes a hill-climbing approach to sparse clustering, which has been shown to be competitive with existing methods in literature on simulated and real-world datasets. The second one considers a semiparametric mixture model for clustering and we propose a semiparametric EM algorithm to fit the model. The third one discusses the difficulty we uncovered in clustering with radial distributions. Under mild conditions, we prove that the magnitudes of the norm of observations sampled from a radial distribution are highly concentrated as the dimension becomes large.

Advisor: Ery Arias-Castro

Monday, May 22, 2017

11:00 AM

AP&M 2402
