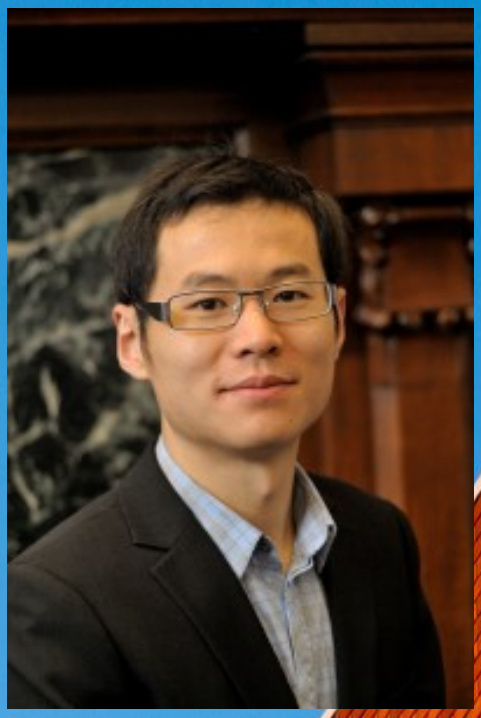
UNIVERSITY OF CALIFORNIA
UC RIVERSIDE

STATISTICS DEPARTMENT SEMINAR

Xiaohui Chen, Ph.D.

Professor

University of Illinois at Urbana-Champaign



Olmsted Hall 420

May 14th 2019

3:45-4:45pm

Reception in Olmsted 1331 at 3:15 P.M.

"Diffusion K-means clustering on manifolds: provable exact recovery via semidefinite relaxations"

FOR MORE INFORMATION ABOUT THIS SEMINAR, VISIT STATISTICS.UCR.EDU/COLLOQUIA.HTML

Abstract

We introduce the diffusion K-means clustering method on Riemannian submanifolds, which maximizes the within-cluster connectedness based on the diffusion distance. The diffusion K-means constructs a random walk on the similarity graph with vertices as data points randomly sampled on the manifolds and edges as similarities given by a kernel that captures the local geometry of manifolds. Thus the diffusion K-means is a multi-scale clustering tool that is suitable for data with non-linear and non-Euclidean geometric features in mixed dimensions. Given the number of clusters, we propose a polynomial-time convex relaxation algorithm via the semidefinite programming (SDP) to solve the diffusion K-means. In addition, we also propose a nuclear norm (i.e., trace norm) regularized SDP that is adaptive to the number of clusters. In both cases, we show that exact recovery of the SDPs for diffusion K-means can be achieved under suitable between-cluster separability and within-cluster connectedness of the submanifolds, which together quantify the hardness of the manifold clustering problem. We further propose the localized diffusion K-means by using the local adaptive bandwidth estimated from the nearest neighbors. We show that exact recovery of the localized diffusion K-means is fully adaptive to the local probability density and geometric structures of the underlying submanifolds.

Biography

Xiaohui Chen is an Assistant Professor of Statistics at the University of Illinois at Urbana-Champaign. He obtained the Ph.D. degree in Electrical and Computer Engineering from the University of British Columbia. His current research interests include: high-dimensional and nonparametric statistics, statistical machine learning and signal processing, and time series analysis. He is a recipient of the NSF CAREER Award (2018).