November 15, 2021 Seminar

Xiaohui Chen from UIUC

Title: A diffusion perspective of manifold clustering

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. 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 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. Joint work with Yun Yang (UIUC).