Optimality of Spectral Clustering for Gaussian Mixture Model

Anderson Ye Zhang

The Wharton School, University of Pennsylvania
E-mail: ayz@wharton.upenn.edu

Abstract: Spectral clustering is one of the most popular algorithms to group high dimensional data. It is easy to implement and computationally efficient. Despite its popularity and successful applications, its theoretical properties have not been fully understood. The spectral clustering algorithm is often used as a consistent initializer for more sophisticated clustering algorithms. However, in this paper, we show that spectral clustering is actually already optimal in the Gaussian Mixture Model, when the number of clusters of is fixed and consistent clustering is possible. Contrary to that spectral gap conditions are widely assumed in literature to analyze spectral clustering, these conditions are not needed in this paper to establish its optimality.