

Boston University Statistics Seminar Series

Modal EM for Mixtures and its Application in Clustering

Surajit Ray

Mathematics and Statistics
Boston University

Thursday, September 28, 2006, 4:00-5:00pm
Mathematics and Computer Science (MCS) Building, Room 149
111 Cummington Street, Boston

Tea and Cookies at 3:30pm in MCS 153

Abstract: Multivariate mixtures provide flexible methods for both fitting and partitioning high-dimensional data. Ray and Lindsay(2005) show that the topography of multivariate mixtures, in the sense of their key features as a density, can be analyzed rigorously in lower dimensions by use of a ridgeline manifold that contains all critical points as well as the ridges of the density. In addition, we have developed a new computing procedure similar to the EM algorithm that can quickly find the modes of a mixture density. This tool can be extended to examine the degree of separation between the modes based on the ridgeline separating them.

These tools can be used in various ways. For one, we can take a conventional mixture analysis and cluster together those components whose contribution is actually unimodal. This cluster could then represent a single true component with a more complex distribution. We can also turn kernel density estimation into a hierarchical clustering tool in which the data points become identified with each other by their association with a common mode of the density estimator. Separate clusters must then correspond to gaps in the estimated density. The analysis is multi-scale, as different levels of smoothing provide different aggregations.

For directions and maps, please see <http://math.bu.edu/research/statistics/statseminar.html>.