

Model-Based Clustering without Parametric Assumptions

David Hunter

Penn State Department of Statistics

This talk discusses finite mixture models in which the component distributions are not assumed to come from any particular parametric family. We begin with a discussion of the essential question of parameter identifiability and introduce an EM-like framework often used to fit these models. We extend these ideas to the multivariate case, which is actually easier in some sense than the univariate case, and introduce the important assumption of conditional independence. We show how to construct an EM-like algorithm, based on a majorization-minimization idea, with desirable theoretical properties. Finally, we extend the multivariate model so that conditional independence need not be assumed. This extension uses a well-developed technique known as independent component analysis (ICA) to create a hybrid estimation algorithm. We illustrate this new methodology using applications in unsupervised learning and image processing, and we discuss what is and is not known about the theoretical properties of the model and the algorithm.