An integrated formulation for merging mixture components based on posterior probabilities

Marc Comas-Cufí   Josep-Antoni Martín-Fernández
Glòria Mateu-Figueras

March 31, 2015

In parametric clustering, groups are usually formed from the components of a finite mixture distribution. The method comprises two steps: first, a finite mixture distribution with the following probability density function

$$f(\,\cdot\,;\pi_1,\ldots,\pi_k,\theta_1\ldots\theta_k) = \pi_1 f(\,\cdot\,;\theta_1) + \ldots + \pi_k f(\,\cdot\,;\theta_k)$$

is fitted to a data set; second, each observation x is assigned to the component j, $1 \leq j \leq k$, with $\hat{\pi}_j f(x;\hat{\theta}_j)$ maximum. After the fitting process, some mixture components might not be separated enough. To deal with this situation, several authors have proposed merging methods that, based on the posterior probabilities incrementally, combine those components that are more similar.

Using a generic definition, an integrated formulation to unify such merging methods is presented and discussed. This new formulation opens the way to define new methods based on the posterior probabilities.