A Doubly-Enhanced EM Algorithm for Model-Based Tensor Clustering

Qing Mai

Florida State University
E-mail: mai@stat.fsu.edu

Abstract: Modern scientific studies often generate tensor data, which calls for innovative statistical analysis methods. An important problem is to perform tensor clustering to understand the heterogeneity in the data. Many existing clustering methods are based on the K-means clustering and ignore the correlation among features. We propose a model-based approach to enable probabilistic interpretation. Our statistical model leverages the tensor structure to reduce the number of parameters for parsimonious modeling. Moreover, our model explicitly exploits the correlation for better variable selection and clustering. We propose a doubly-enhanced EM (DEEM) algorithm to perform clustering under this model. Both the E-step and the M-step are carefully tailored for tensor data. Theoretical studies confirm that DEEM achieves consistent clustering even when the dimension of each mode of the tensors grow at an exponential rate of the sample size, while numerical studies demonstrate favorable performance of DEEM in comparison to existing methods.