A penalized BIC for mixture model selection

Orla Murphy
Dalhousie University

Abstract

The Bayesian information criterion (BIC) is the most popular criterion for mixture model selection. In high-dimensional settings, a key drawback of this criterion is its tendency to underestimate the number of mixture components. This work focuses on model selection for Gaussian mixtures and, in particular, mixtures of factor analyzers (MFA). In MFA, a p-dimensional random vector is replaced with $q < p$ latent factors thereby reducing the number of free parameters in the covariance structure from quadratic to linear in p. For this model, the number of latent factors must be estimated in addition to the usual estimation of the number of components. A penalized likelihood is introduced to further reduce the number of free parameters in the MFA model, and a penalized BIC is used for model selection. Different from previous work, a penalty applies to the component covariance matrices through the factor loading matrices. Important properties of this criterion will also be discussed.