

Model-Based Clustering of Longitudinal Data

Paul David McNicholas¹ and Thomas Brendan Murphy²

¹ Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada.

² School of Mathematical Sciences, University College Dublin, Dublin, Ireland.

Model-based clustering is a method of clustering data based on mixture modeling. Within the popular literature, a Gaussian mixture model is most frequently used. A common approach to model-based clustering exploits an eigenvalue decomposition of the group covariance matrices to give a wide range of covariance structures. Other approaches include the mixtures of factor analyzers and mixtures of probabilistic principal components analyzers models. Model parameters are typically estimated using an expectation-conditional maximization (ECM) algorithm or an alternating expectation-conditional maximization (AECM) algorithm.

These ideas around model-based clustering are extended to facilitate application to longitudinal, or time-ordered, data *via* a modified Cholesky decomposition of the covariance matrices. For each mixture component, a variety of constraints are imposed to give a family of eight mixture models. These models are fitted using an ECM algorithm. This new family of mixture models is applied to data on the weight of rats on one of three different diets, where it gives excellent cluster-capturing performance.