Mixture Models, Clustering and Classification
Chair: Abbas Khalili (McGill University)

LIAM CALLAGHAN, University of Guelph
Modelling Plant-Pollinator Interactions with Mixtures of Linkage Rules

In a given ecosystem, there may be several mechanisms facilitating an observed interaction between a plant and pollinator. Hence pollination networks should be modelled as a mixture of linkage rules. We propose Latent Dirichlet Allocation (LDA) from artificial intelligence to model the observed interactions in an ecosystem as a finite mixture of (latent) interaction groups in which plant and pollinator pairs with common linkage rule(s) are placed in an interaction group. LDA is studied through a simulation study using the BIC for model selection. The results suggest that LDA works well for compartmental webs, but loses accuracy as nestedness increases.

ARMIN HATEFI, University of Manitoba
Fisher Information in Ranked Set Samples from Finite Mixture Models

We study the efficiency of ranked set sampling (RSS) relative to simple random sampling (SRS) for making inference about the parameters of a finite mixture model using the Fisher information matrix. We propose different variations of RSS (for both perfect and imperfect ranking mechanisms) and show how to obtain the Fisher information matrix for each variation. Our findings show that statistical inference about different features of finite mixture models based on RSS techniques is more efficient than the corresponding ones using SRS. Theoretical results are augmented by a simulation study based on a mixture of two univariate normal distributions.

POOYAN KHAJEHPOUR TADAVANI, University of Waterloo
Partitioning High Dimensional Data to Low Dimensional and Localized Subspaces

It is generally conceivable to assume that high dimensional data points are on submanifolds of the space. These submanifolds can be modeled by a number of linear subspaces. This is the main intuition behind a majority of subspace clustering algorithms. However, subspaces computed by these algorithms, consist of disconnected subsets of the submanifolds and therefore, they do not form localized clusters. We propose Low Dimensional Localized Clustering (LDLC), a new method for subspace clustering. LDLC, unlike existing methods, respects the topology of the submanifolds and assigns the data points to localized clusters such that the total reconstruction error is minimized.

IRENE VRBIK, University of Guelph
Clustering and Classification with Parsimonious Skew-t Mixture Models

With the advancement of computer technology, mixture model-based approaches to clustering have become increasingly popular. In recent work, a robust, flexible mixture modelling approach using the skew-t distribution has been explored. We propose a skew-t analogue of the popular MCLUST models that impose an eigenvalue decomposition of the covariance structure. An “exact” EM algorithm is outlined and our approach is applied to some benchmark clustering datasets.

BRIAN FRANCZAK, University of Guelph
The ParSAL Family of Mixture Models

We introduce a family of shifted asymmetric Laplace (SAL) distributions for model-based clustering and classification. These models will arise through eigen-decomposition of the component covariance matrices. An expectation-maximization (EM) algorithm is used for parameter estimation and the integrated completed likelihood (ICL) for model selection. The capabilities
of our parsimonious SAL family are demonstrated on both real and simulated data and the results compared to the popular Gaussian alternative.

RACHEL O’REILLY, University of Guelph
Cross-Validation for Mixture Model Selection

In recent years, the use of mixture models for classification and clustering has gained popularity. Although there are existing methods for model selection, such as the BIC, AIC, and ICL, no one approach has proven particularly effective. We investigate the use of cross-validation to assess the consistency of classification under a potential model, and thus assess the certainty of the model in revealing underlying data structure. The procedure will be illustrated using both real and simulated data.