ALPEN-ADRIA UNIVERSITÄT KLAGENFURT

INVITATION TO THE DEFENSE

Bayesian methods for model-based clustering

Alexander Mozdzen | University of Klagenfurt

Chaired by Gregor Kastner | Department of Statistics

Friday, 10 January 2025 | 09:30–11:30

S.1.05

ABSTRACT

This research advances Bayesian methods for model-based clustering, focusing on innovative applications and methodological developments in mixture models. Mixture models, widely used in Bayesian statistics for clustering and density estimation, can be specified to have a fixed, random, or infinite number of components. The flexibility to model an infinite number of components is achieved through the Dirichlet process (DP), a cornerstone of Bayesian nonparametrics. By strategically placing the DP prior on selected model parameters, it becomes possible to tailor clustering frameworks to address specific research questions.

A key focus of this work is the application of the DP to regression models. One such model is a spatiotemporal linear regression framework designed to analyze unemployment rates in Italy, where the DP is used to cluster areal units based on their economic characteristics. By leveraging the sparsity properties of Gaussian Markov random fields, an efficient Markov chain Monte Carlo (MCMC) algorithm was developed. This approach not only demonstrated competitive out-of-sample performance but also provided insights into the economic disparities across Italian provinces.

Extending the application of the DP, a multinomial logit regression model was developed to examine the impact of agricultural subsidies on land use across Europe. This framework introduces a partial clustering approach, where the DP prior is applied selectively to a subset of regression coefficients, allowing clustering to be informed by environmental variables while controlling for economic and physical characteristics. Using advanced data augmentation techniques, the model identified distinct clusters in the data, shedding light on the heterogeneous effects of subsidies on land use.

From a methodological perspective, this research introduces a novel prior for mixture weights in clustering. A persistent challenge in mixture models is the tendency to estimate redundant or overlapping clusters. To address this, a repulsive property was incorporated into the prior on mixture weights, favoring dissimilar weights and promoting more parsimonious solutions. The theoretical properties of this prior were analyzed, and its performance was validated in a simulation study, where it demonstrated an improved ability to resolve similar cluster weights and determine the optimal number of clusters. The approach was further applied to a biomedical dataset, clustering children based on body mass index and eating behavior, resulting in interpretable and meaningful groupings.

We hope to see you there!