Editorial

Published online: 2 April 2010
© Springer-Verlag 2010

This first issue of volume 4(2010) of the journal *Advances in Data Analysis and Classification* (ADAC) contains four articles. They deal with ridge regression clustering, Gaussian mixtures, the comparison of partitions, and with the selection of tag SNPs in micro-array analysis.

The paper by *Christian Hennig* on ‘Methods for merging Gaussian mixture components’ considers the situation when, for a given data set, a Gaussian mixture has been fitted, but the obtained mixture components are not sufficiently separated in order to be interpreted as ‘clusters’. Then some components must be merged. The paper shows various methods for defining suitable cluster concepts (based on unimodality and misclassification probabilities) and corresponding merging strategies. The separation of mixture components is visualized by an ‘ordered posterior plot’ and the method is illustrated by data examples.

The second paper starts from the remark that classical regression clustering methods, as well as their fuzzy counterparts, may suffer from eventual multicollinearity among predictor variables. The paper by *Hye Won Suk* and *Heungsun Hwang* on ‘Regularized fuzzy clusterwise ridge regression’ proposes a new approach to cope with this situation. It combines ridge regression (for the estimation of cluster-specific regression coefficients) with regularized fuzzy clustering based on maximization of entropy (for determining an optimal degree of fuzziness). The usefulness of the approach is demonstrated by simulated and real-data examples (e.g., considering characteristics of used cars).

In clustering contexts and applications there are situations where the same variables (e.g., the items from a questionnaire) are recorded for two different sets of units (e.g., the members of two teams) and a given partition of the first set is to be compared with a partition of the second set on the basis of the two recorded data matrices for the same variables. In their paper ‘Comparing partitions of two sets of units based on the same variables’ *Genane Youness* and *Gilbert Saporta* propose a new approach for solving this comparison problem. Their approach combines discriminant methods, the consideration of various association indices (under the hypothesis of identical

partitions), and simulated random partitions within a common latent class model (in order to estimate the probability distribution of the indices). Simulated and real-data examples illustrate their approach.

Finally, the paper ‘Tag SNP selection based on clustering according to dominant sets found using replicator dynamics’ by *Florian Frommlet* considers the problem of finding those genetic regions (markers) on a DNA strain that are associated with a disease (diabetes, cancer type) or a quantitative trait (blood pressure). More specifically, he looks at single nucleotide polymorphisms (SNPs, single positions of DNA with variation among individuals) that may have causal relationship to the trait. The paper explains the genetic background and also how associations (LD, linkage disequilibrium) between SNPs can be measured from data obtained by SNP-array experiments. These measures define an association graph among SNPs and it is common practice to define, and search, ‘important’ SNPs (tag SNPs) within this graph by graph-theoretical concepts, partitioning or covering problems, and corresponding algorithms. The current paper adds to this toolbox a new method that proceeds essentially by covering the graph by disjoint cliques. It uses a clustering strategy based on the concept of dominant sets and proposes a suitable clustering algorithm. Tag SNPs are then selected as representatives of the obtained clusters. In a series of simulated examples (based on 30 real data sets chosen from the HapMap database) the new method is compared with Tagger, one of the most popular SNP selection methods. Clustering provides never worse, but often quite better statistical power than Tagger.

Hans-Hermann Bock (Aachen)
Wolfgang Gaul (Karlsruhe)
Akinori Okada (Tokyo)
Maurizio Vichi (Rome)