

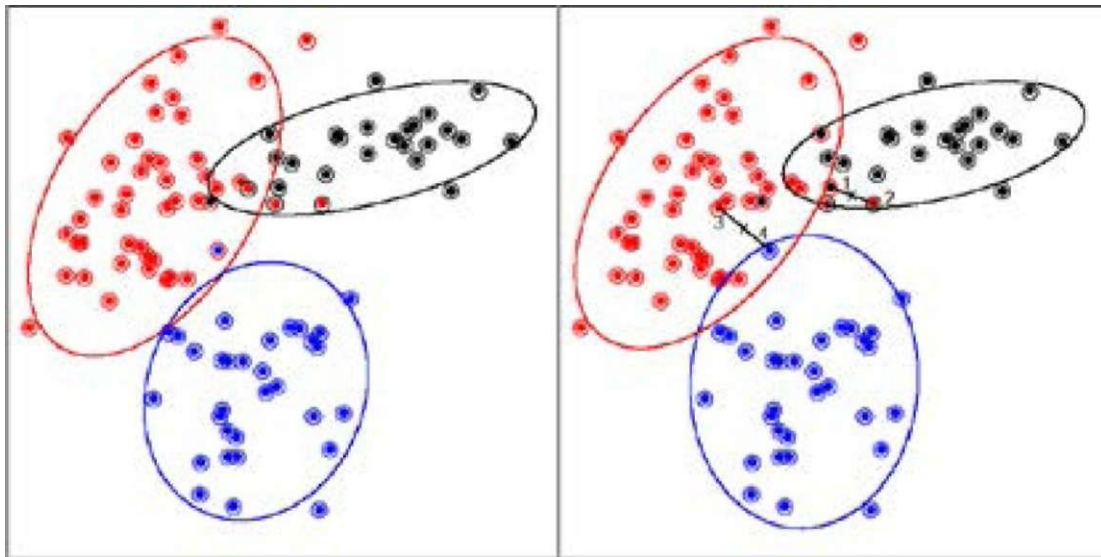
## THE USE OF POSITIVE AND NEGATIVE EQUIVALENCE CONSTRAINTS IN MODEL-BASED CLUSTERING

V. Melnykov<sup>1</sup>, I. Melnykov<sup>2</sup>, S. Michael<sup>1</sup>

1) Department of Information Systems, Statistics, and Management Science, University of Alabama, USA; \*[vmelnykov@ua.edu](mailto:vmelnykov@ua.edu); 2) Department of Mathematics, School of Science and Technology, Nazarbayev University, Kazakhstan.

**Introduction.** Cluster analysis is a popular technique in statistics and computer science with the objective to group similar observations into relatively distinct groups known as clusters. Semi-supervised model-based clustering assumes that some additional information about group memberships is available.

**Methodology.** We discuss a general type of semi-supervised clustering defined by positive and negative equivalence constraints. Positive constraints require of some points to belong to the same cluster. To the contrary, negative constraints specify which points are not allowed to be in the same cluster. Such restrictions can lead to a model that describes the specifics of the data more accurately as shown in Figure 1 where negative equivalence constraints were applied. The EM algorithm technique was used for fitting a finite mixture model [1].



*Fig. 1.* Unsupervised (left) and semi-supervised (right) clustering of a simulated dataset.

**Results and discussion.** Positive equivalence constraints are more easily accommodated by the EM algorithm and allow for more tractable analytic expressions [2]. At the same time, the implementation of negative constraints depends on their specific configuration and number of blocks involved. We have developed a universal technique for the inclusion of both kinds of restrictions in the E-step of the EM algorithm. Our approach shows good results on simulated and real-life datasets. The benefits of the proposed method result from the ability of the algorithm to work over a restricted parameter space, thus resulting in faster convergence to the solution given that thorough initialization was performed.

### References.

1. McLachlan, G. and Peel, D. (2000), *Finite Mixture Models*, New York: John Wiley and Sons, Inc.
2. Shental, N. et al. (2003), Computing Gaussian mixture models with EM using equivalence constraints, in *Advances in NIPS*, vol. 15.