Model based clustering techniques

1. Introduction to model-based clustering concepts and general framework (2 sessions)

    - What is cluster analysis? (Supervised and unsupervised learning).
    - Types of cluster analysis algorithms.
    - Similarity and dissimilarity measures in clustering.
    - Definition of model-based clustering.
    - Measures for evaluation of the quality of clustering.

2. Univariate and multivariate mixture distributions: theory and practice with R (4 sessions + 2 sessions practical)

    - Definition of finite mixture model (FMM).
    - Univariate and multivariate Gaussian mixture model (GMM).
    - Estimation parameters in GMM.
    - Expectation Maximization (EM) algorithm.
    - Selection number of components in mixture models.
    - Clustering real and simulation data with R, ("cluster" and "mclust" packages)

3. Introduction to α-stable distributions and sub-Gaussian α-stable (SGαS) mixture model (4 sessions).

    - Definition of α-stable distributions.
    - Examples of using α-stable distributions in real world.
    - Sub-Gaussian alpha-stable mixture model (SGαSMM).
    - Pseudo-stochastic EM for SGαSMM.
    - Cluster-weight model based on SGαS.
    - SGαSMM in R (stable package).

4. Spatial Clustering (4 sessions)

    - Introduction to spatial data
    - k-means algorithm for spatial data
    - Hierarchical algorithms for clustering spatial data

- Density-Based Spatial Clustering
- Spatial clustering in R

5. Bayesian approach: Introduction and developments (8 sessions)

- Bayesian Inference
- Bayesian Inference for a Mixture Model using the Gibbs Sampler
- Bayesian model based clustering procedures

    1: The prior part of the clustering model

    2: The data part of the clustering model
- High-Dimensional Bayesian Clustering with Variable Selection: The **R** Package **bclust**

During these seminars students can know with "cluster analysis" and its applications in real life. Here, our attention is focus on using statistical methods for clustering data. A variety of data with and without outliers are used in our models to illustrate applications of presented algorithms. In spatial clustering, we cluster dependent data, which is useful in some real phenomena like **geostatistics** and **environmental** science. The Bayesian inference of mixture models has many advantages over the classical approach and can increase precision of clustering. However, it is not without computational difficulties. We, present some more popular of Bayesian clustering methods, and students learn how they can use them in practice to cluster high-dimensional data.

During the class, students will learn advanced theory of model-based cluster analysis. Case studies will be developed during practical sessions to introduce Ph.D. students to advanced applications in the field.

The course will also address the clustering of dependent variables (spatial clustering) that are found in real life application (environmental studies, climatological analysis etc.). Computational difficulties will be given considerable attention so that students will develop implementation capacities.

Furthermore, robust model-based clustering based on sub-Gaussian stable distribution are taught. Comparison of the classical and Bayesian approaches will stimulate the critical capability of the class. Students will be encouraged to develop "statistical thinking" at an advanced level.

Group work will be encouraged to develop case studies in the practical sessions.

The students are expected to develop collaborative work skills in research subjects.