



# A relational approach in classification, subgroup discovery and statistical data analysis

Georg Schollmeyer  
(Institut für Statistik, LMU)

13.02.2019, 16.00 (s.t.)  
Institut für Statistik, Ludwigstr. 33, Seminarraum

Many methods in statistical data analysis and also in classification make use of notions of proximity of data points, mostly formalized by conceptually embedding the data into a metric space or a space equipped with more structure (e.g. a Hilbert space). In this talk, I want to present an approach that does not emphasize the notion of distance, but instead uses relational notions of location like that of incidence or betweenness to describe and analyze the data. Given a data set with a binary target variable of interest, in the spirit of subgroup discovery, this relational approach looks at subgroups that are very interesting w.r.t. the target variable. In contrast to classical subgroup discovery, we do not look at subgroups that are described explicitly by characteristics of certain attributes. Instead, we look at subgroups that are described implicitly by notions of incidence and betweenness. The approach can be used both for more classical statistical tasks like e.g. testing the homogeneity of the distribution of the target variable within subgroups, as well as for an approach to classification which is very much in the spirit of a relational ‘counterpart’ to the idea of nearest neighbors. After introducing the basic ideas, I would like to discuss statistical properties of the approach, computational aspects and especially the problem of data with different scales of measurement, the problem of high dimensional data, as well as a concrete proposal for regularization.