Variable Selection by Regularization Methods for Generalized Mixed Models

Andreas Groll[1]

Department of Statistics, University of Munich, Akademiestrasse 1, D-80799, Munich, Germany
(E-mail: andreas.groll@stat.uni-muenchen.de)

Rigorosumsvortrag
1. Dezember 2011, 13:30 Uhr, **Seminarraum**, Ludwigstrasse 33 I

Abstract. Generalized mixed models are a widely used tool for modeling longitudinal data. However, their use is typically restricted to few covariates, because the presence of many predictors yields unstable estimates. In the talk regularization techniques for generalized mixed models are developed that are able to perform variable selection. These techniques are especially appropriate when many potential influence variables are present and existing approaches tend to fail.

Boosting methods for different generalized mixed models are presented, which are based on the likelihood function and work by iterative, componentwise fitting of the residuals using weak learners with an implicit selection of relevant variables. The complexity of the resulting estimators is determined by information-based criteria. Boosting originates in the machine learning community. It represents a general method for improving the accuracy for any given learning algorithm. Likelihood-based boosting may be seen as an extension of L2 boost.

Furthermore, an L_1-penalization technique for generalized linear mixed models is presented, which enforces variable selection and shrinkage simultaneously. A gradient ascent algorithm is proposed that allows to maximize the penalized log-likelihood yielding models with reduced complexity. In contrast to common procedures it can be used in high-dimensional settings where a large number of potentially influential explanatory variables is available.

The methods are investigated in simulation studies and illustrated by use of a real data set.
Keywords: Generalized mixed model, Boosting, Penalized Quasi-Likelihood, Laplace approximation, Lasso, Gradient ascent, Penalty, Variable selection.