



Approaches for sharing information between heterogeneous patient subgroups in sparse Cox models

Katrin Madjar

(Department of Statistics, LMU Munich)

27.11.2019, 16.00 (s.t.)

Department of Statistics, Ludwigstr. 33, Room 144

The construction of risk prediction models for specific subgroups of patients based on high-dimensional molecular measurements such as gene expression data is an important task in clinical medicine. Main objectives in modeling high-dimensional data are good prediction performance and feature selection to find a subset of predictors that are truly associated with a clinical outcome such as time-to-event endpoint. This task is often challenging in practice since patient cohorts are typically small and can be heterogeneous with regard to their relationship between predictors and outcome. It is tempting to combine patient subgroups to increase sample size. However, due to the heterogeneity between subgroups results can be biased and subgroup-specific effects may get lost.

For this situation, we propose two different approaches, a frequentist Cox model and a Bayesian Cox model with gene expression data as covariates and survival outcome. We aim at providing a separate prediction model for each subgroup that allows for identifying common as well as subgroup-specific effects and has improved prediction accuracy over standard approaches. The frequentist Cox model uses a lasso penalty for variable selection and a weighted version of the partial likelihood that includes patients of all subgroups but assigns them individual weights based on their subgroup affiliation. Patients who fit well to the subgroup of interest receive higher weights in the subgroup-specific model. For the Bayesian model we assume a network linking genes within and across different subgroups. Network information is incorporated into variable selection to help identifying pathways of functionally related genes and genes that are simultaneously prognostic in different subgroups. Both approaches are applied to simulated data and real lung cancer cohorts.