Comparison of approaches for fitting generalized additive models

Harald Binder
Institut für Medizinische Biometrie und Medizinische Informatik,
Universitätsklinikum Freiburg, Germany

Generalized additive models are popular when a regression model is to be fitted to a non-normal response where the influence of single covariates is assumed to have unknown non-linear shape. One important difference of the procedures available within the R environment for fitting such models is how covariates and effective degrees of freedom for each covariate are selected. We evaluate the optimization approach of the recommended R package `mgcv`, a stepwise approach, and a mixed model approach. For comparison we offer a new fitting procedure `GAMBoost` – based on boosting techniques – which is built to perform implicit selection of covariates for high-dimensional problems. Its implementation is made available as a new R package. For comparison we focus on simulated data with a small signal-to-noise ratio and/or a large number of covariates. For the underlying true structure simple linear models as well as models incorporating non-linear covariate effects are used. The former allow for comparison of the R packages in situations where the class of models offers too much flexibility while the latter require fitting of complex structure with a limited amount of data. Performance of the procedures is evaluated with respect to prediction performance as well as with respect to the identification of influential covariates. In addition settings are identified where many procedures do not to return any fit at all and only `GAMBoost` provides a viable alternative.