

Semiparametric Mixed Models and Boosting

Gerhard Tutz

Ludwig-Maximilians-Universität München, Institut für Statistik

tutz@stat.uni-muenchen.de

In linear mixed models which are one of the most widely used tools in applied statistics the influence of covariates is restricted to a strictly parametric form. With the rise of semi- and nonparametric regression also the mixed model has been expanded to allow for more flexible forms of the predictor, Lin and Zhang (1999) propose additive models based on smoothing splines whereas Wand (2003) uses basis functions with penalization to model the smooth term. The proposed alternative approach is likelihood based boosting. Boosting originates in the machine learning community where it has been proposed as a technique to improve classification procedures by combining estimates with reweighted observations (e.g. Freund, 1995). Likelihood based boosting is a general method which may be seen as an extension of L2 boost as considered by Bühlmann and Yu (2003) in regression modelling. The advantage of boosting techniques in the form of componentwise boosting is that it is suitable for high dimensional settings where many influence variables are present. It allows to fit additive models for many covariates with implicit selection of relevant variables and automatic selection of smoothing parameters. Moreover, boosting techniques may be used to incorporate monotonicity restrictions on the additive terms as well as to fit smooth random effects distributions which are non-Gaussian.

References:

Bühlmann, P. and Yu, B. (2003). Boosting with l_2 loss: Regression and classification. *Journal of the American Statistical Association* **98**, 324–339.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation* 121, 256–285.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society* **B61**, 381–400.

Wand, M. P. (2003). Smoothed and mixed models. *Computational Statistics* **18**, 223–249.