Peeking into the black box: recursive partitioning of (generalized) linear models

Thomas Rusch[1] **and Achim Zeileis**[2]

[1] *Vienna University of Economics and Business, Austria*
[2] *Innsbruck University, Austria*

Abstract

Recursive partitioning algorithms separate a feature space into a set of disjoint rectangles. Typically, a constant (e.g., a mean or a proportion) is fitted in every segment of the partition. While this is a simple and intuitive approach, it still lacks interpretability as to how a specific relationship between dependent and independent variables may look. Or it may be that a certain model is assumed or of interest and there is a number of candidate variables that may nonlinearily give rise to different model parameter values. We want to present an approach that offers a solution to the problem of limited interpretability of classical trees as well as providing an explorative way to assess a candidate variable's influence on a parametric model: Model-Based Recursive Partitioning (Zeileis et.al., 2008). This method conducts recursive partitioning of a parameteric model such as the generalized linear model by (1) fitting a parametric model to a data set, (2) testing for parameter instability over a set of partitioning variables, (3) splitting the model with respect to the variable associated with the highest instability. The outcome is a tree where each node is associated with a fitted parametric model. We will describe the procedure and show its versatility and suitability to gain additional insight into the relationship of dependent and independent variables by three examples, the link between professors' beauty and their teaching evaluation, the prediction of voting behaviour and a failure model for debt amortization.

Keywords

Model-based recursive partitioning, Trees, Generalized linear model, Parameter instability, Maximum likelihood.

References

Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist. 17*, 492–514.