DSSV 2023
Data Science, Statistics & Visualisation

EC DA
2023
EUROPEAN
CONFERENCE ON
DATA ANALYSIS

Optimization of Decision Trees under Restrictions: The PrInDT package

Claus Weihs[1] and Sarah Buschfeld[1]

[1]TU Dortmund University, Dortmund, Germany

PrInDT is an R-package [1] for the optimization of conditional inference trees (ctrees) for classification and regression. For optimization, the model space is searched for the best tree on the full sample by means of repeated subsampling. Restrictions are allowed so that only trees are accepted which do not include pre-specified uninterpretable split results. With the PrInDT package, both the predictive power and the interpretability of ctrees are increased. The performance of ensembles and individual trees is compared.

The package covers the optimization of ctrees for 2-level, multilevel, and multilabel classification as well as for regression. Subsampling percentages can be varied for the classes in classification and for observations and predictors in regression. Finally, the posterior distribution of a specified variable in the terminal nodes of a given tree can be analyzed.

The usage of the functions is illustrated by means of examples from linguistics.

[1] Claus Weihs, Sarah Buschfeld (2023): PrInDT: Prediction and Interpretation in Decision Trees for Classification and Regression, R package version 1.0, url = https://CRAN.R-project.org/package=PrInDT