Granered, Nicholas; Prins, Samantha C. Bates
Identifying outlying observations in regression trees. (English) Zbl 1348.62134

Summary: Regression trees are an alternative to classical linear regression models that seek to fit a piecewise linear model to data. The structure of regression trees makes them well-suited to the modeling of data containing outliers. We propose an algorithm that takes advantage of this feature in order to automatically detect outliers. This new algorithm performs well on the four test datasets [A. Hadi and J. Simonoff, “Procedures for the identification of multiple outliers in linear models”, J. Am. Stat. Assoc. 88, No. 424, 1264–1272 (1993; doi:10.1080/01621459.1993.10476407)] that are considered to be necessary for a valid outlier detection algorithm in a linear regression context, even though regression trees lack the global linearity assumption. We also show the practical use of this approach in detecting outliers in an ecological dataset collected in the Shenandoah Valley.

MSC:
62G08 Nonparametric regression and quantile regression

Keywords:
outlier detection; influential observations; backward-stepping; robust models; outlier; CART

Full Text: Euclid