Bar-Hen, Avner; Gey, Servane; Poggi, Jean-Michel

Influence measures for CART classification trees. (English) Zbl 1331.62300

Summary: This paper deals with measuring the influence of observations on the results obtained with CART classification trees. To define the influence of individuals on the analysis, we use influence measures to propose criterions to quantify the sensitivity of the CART classification tree analysis. The proposals are based on predictions and use jackknife trees. The analysis is extended to the pruned sequences of CART trees to produce CART specific notions of influence. Using the framework of influence functions, distributional results are derived.

A numerical example, the well known spam dataset, is presented to illustrate the notions developed throughout the paper. A real dataset relating the administrative classification of cities surrounding Paris, France, to the characteristics of their tax revenues distribution, is finally analyzed using the new influence-based tools.

MSC:
62H30 Classification and discrimination; cluster analysis (statistical aspects)
62G35 Nonparametric robustness
62P20 Applications of statistics to economics

Keywords: influential individuals; influence functions; decision trees; CART

Software:
R; MASS (R); ElemStatLearn

Full Text: DOI Link

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.