Garrote trees as tree structured regression analysis.

Summary: In regression analysis, stochastic models are often constructed to model relationships between outcomes and explanatory variables. We derive statistical interpretation about the underlying structure of data based on these models. When we use a linear regression model and the model provides good fitting to the data, it is straightforward to interpret the relation. However, there are cases where it may be difficult to formulate a linear model reflecting actual characteristics in detail. In such cases, a tree-structured approach is recommended, such as classification and regression trees (CART), which develops a tree and provides an interpretation of the data based on the fundamental model derived from the tree. Random forest (RF) involves an ensemble learning method based on the trees and can predict outcomes more precisely. However, RF cannot provide a tree-structured model for interpreting the data. We examine a nonnegative garrote (NNG), a shrinkage estimator, and propose garrote trees (GT) as an adjustment of RF based on NNG. In addition, GT can lead making trees that are useful for interpretation of data. Two case studies of diabetes and prostate cancer data illustrate predictive accuracy and descriptive features of GT. Finally, our simulation studies show that the proposed method is highly accurate predictively and provides a potential ability to interpret the data from new meaningful standpoints.

MSC:

62J05 Linear regression; mixed models
62J07 Ridge regression; shrinkage estimators (Lasso)
62H30 Classification and discrimination; cluster analysis (statistical aspects)
68T05 Learning and adaptive systems in artificial intelligence

Keywords:
tree-structured approaches; ensemble learning; CART; random forest; nonnegative garrote; classification

Full Text: DOI

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.