

Tree-Based Models for Fitting Stratified Linear Regression Models

William Shannon (Washington University in St. Louis School of Medicine)

`shannon@ilya.wustl.edu`

Maciej Faifer, Michael A. Province, and D. C. Rao

Abstract

This paper generalizes the methods developed in Shannon, Province, and Rao (2001) to use recursive partitioning to identify subsets of the aggregate data within each of which simple linear regression models give better fit. This method is proposed as an alternative to multivariate regression modeling when the analyst is primarily concerned with the regression of an outcome onto a single predictor and needs to control for other covariates. Splitting rules and pruning methods are derived, programmed in C, and linked to the public domain 'RPART' software providing a full software implementation of this methodology. Examples are presented to illustrate the methodology and software.