Reynolds, Evan L.; Callaghan, Brian C.; Gaies, Michael; Banerjee, Mousumi
Regression trees and ensemble for multivariate outcomes. (English) Zbl 07683113
Sankhyā, Ser. B 85, No. 1, 77-109 (2023)

Summary: Tree-based methods have become one of the most flexible, intuitive, and powerful analytic tools for exploring complex data structures. The best documented, and arguably most popular uses of tree-based methods are in biomedical research, where multivariate outcomes occur commonly (e.g., diastolic and systolic blood pressure and nerve conduction measures in studies of neuropathy). Existing tree-based methods for multivariate outcomes do not appropriately take into account the correlation that exists in such data. In this paper, we develop goodness-of-split measures for building multivariate regression trees for continuous multivariate outcomes. We propose two general approaches: minimizing within-node homogeneity and maximizing between-node separation. Within-node homogeneity is measured using the average Mahalanobis distance and the determinant of the variance-covariance matrix. Between-node separation is measured using the Mahalanobis distance, Euclidean distance and standardized Euclidean distance. To enhance prediction accuracy we extend the single multivariate regression tree to an ensemble of multivariate trees. Extensive simulations are presented to examine the properties of our goodness-of-split measures. Finally, the proposed methods are illustrated using two clinical datasets of neuropathy and pediatric cardiac surgery.

MSC:
62H30 Classification and discrimination; cluster analysis (statistical aspects)
62P10 Applications of statistics to biology and medical sciences; meta analysis
68W01 General topics in the theory of algorithms

Keywords:
multivariate outcomes; regression trees; Mahalanobis distance; clinical interpretability; machine learning

Full Text: DOI

References:
[8] Callaghan, BC; Xia, R.; Banerjee, M.; de Riekeneire, N.; Harris, TB; Satterfield, S.; Schwartz, AV; Vinik, AI; Feldman, EL; Strotmeyer, ES, Metabolic syndrome components are associated with symptomatic polyneuropathy independent of glycemic status, Diabetes Care, 39, 801-807 (2016) · doi:10.2337/dc16-0081