Customizing the rpart library for multivariate gaussian outcomes: the longRPart library

Sam Stewart[1*], **Mohamed Abdolell**[2], **Michael LeBlanc**[3]

1. Faculties of Medicine and Computer Science, Dalhousie University
2. Department of Diagnostic Radiology, Dalhousie University
3. Fred Hutchinson Cancer Research Center, Seatle, Washington

* sam.stewart@dal.ca

Keywords: repeated measures outcomes, classification trees, rpart

Abdolell et al. (2001)[1] implemented a binary partitioning algorithm for the case of continuous repeated measures outcomes, using Mahalanobis distance as a deviance measure to evaluate goodness-of-split. The algorithm was implemented only for a single split at the root node of the tree with the single purpose of dichotomising prognostic variables. The binary partitioning algorithm was implemented in SAS using the PROC MIXED procedure, along with a permutation test to evaluate the p-value of the associated binary split and a bootstrap method to calculate a confidence interval.

This project extends the binary partitioning algorithm of Abdolell et. al to a binary recursive partitioning algorithm[2] which is implemented in R. We utilize the *nlme* library to extend the *rpart* library[3], producing the *longRPart* library for binary recursive partitioning in the case of MVN outcomes, and extends the algorithm to split on unordered categorical variables. A tree plotting function is developed for annotated plots that are applied to terminal nodes of the tree to display the longitudinal profiles of the outcome variable.

A detailed discussion will be presented of how the *rpart* library was extended to accomodate the longitudinal outcome with its associated deviance measure, and how to apply these same principles to the case of other non-standard outcomes using custom R functions.

References

Abdolell et al (2002). Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. *Statistics in Medicine* 21(22), 3395-03409.

Breiman, Friedman, Olshen, and Stone. (1984) Classification and Regression Trees. Wadsworth.

Terry M Therneau and Beth Atkinson. R port by Brian Ripley. (2008). rpart: Recursive Partitioning. R package version 3.1-42.
http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm