

Learning Treed Generalized Linear Models

Hugh Chipman (University of Waterloo)
hachipma@icarus.math.uwaterloo.ca

Abstract

Tree models can be an effective and interpretable tool for supervised learning problems (i.e., regression and classification). A recent variation on trees is the “treed model”, which includes a more sophisticated model in each terminal node of the tree, such as a linear regression. This talk considers generalized linear models as a broader class of terminal node models. Specific examples include binary and Poisson regression. One of the main challenges in this area is effective algorithms for learning these models. This talk will explore a Bayesian approach which offers several advantages, including regularization through careful specification of prior distributions, a stochastic search in the tree space, and the potential to improve predictions by model averaging. Data mining applications in areas such as marketing, insurance, and drug discovery will be discussed. Connections with other methods such as Boosting and Bagging will also be explored.