

Regression Trees for Multivariate Count Data

史玉山¹

¹ 中正大學

Abstract

We propose a regression tree method for multiresponse count data. At each node, a univariate count regression model is fitted and its associated partial score residuals are obtained. We then use the conditional independence tests of analysis of contingency tables of the residual patterns to select the split variable. A likelihood approach is applied to the selection of the split point at each node and to the tree pruning process. Our tree method is shown to be free from selection bias. Its advantage in prediction is demonstrated through some simulations and a real data analysis on the demand for medical care.

Keyword: Conditional independence test , Hurdle model, Partial Score residual, Zero-inflated model