

Grouped variable selection in the Cox model

Bin Nan

Department of Biostatistics

School of Public Health

University of Michigan

Ann Arbor, MI 48109-2029

bnan@umich.edu

Abstract

In many biological and other scientific applications, predictors are often naturally grouped. For example, in biological applications, assayed genes or proteins are grouped by biological roles or biological pathways. When studying the dependence of survival outcome on these grouped predictors, it is desirable to select variables at both the group level and the within-group level. We developed a new method to address the group variable selection problem in the Cox proportional hazards model. Our method not only effectively removes unimportant groups, but also maintains the flexibility of selecting variables within the identified groups. We also show that the new method offers the potential for achieving the asymptotic oracle property.