

Resampling-based information criteria for best-subset regression

Philip T. Reiss · Lei Huang ·
Joseph E. Cavanaugh · Amy Krain Roy

Received: 5 July 2010 / Revised: 16 August 2011 / Published online: 20 March 2012
© The Institute of Statistical Mathematics, Tokyo 2012

Abstract When a linear model is chosen by searching for the best subset among a set of candidate predictors, a fixed penalty such as that imposed by the Akaike information criterion may penalize model complexity inadequately, leading to biased model selection. We study resampling-based information criteria that aim to overcome this problem through improved estimation of the effective model dimension. The first proposed approach builds upon previous work on bootstrap-based model selection. We then propose a more novel approach based on cross-validation. Simulations and analyses of a functional neuroimaging data set illustrate the strong performance of our resampling-based methods, which are implemented in a new R package.

Keywords Adaptive model selection · Covariance inflation criterion · Cross-validation · Extended information criterion · Functional connectivity · Overoptimism