Stepwise Paring Down Variation for Identifying Influential Multifactor Interactions

黃景祥

中央研究院統計科學研究所

Abstract

We consider the problem of identifying influential sets of factors related to a continuous response variable from a large number of factor variables. The ability of most available methods to reveal more true factors and fewer false ones often relies heavily on tuning parameters which is a difficult task. We provide a completely different solution with a simple and novel idea for the identification of influential sets of variables. The method is simple as it involves only repeatedly implementing single-term analysis of variation. The main idea is to stepwise pare down the total variation of responses so that the remaining influential sets of factors have increased chances of being identified. We have developed an R package and demonstrated its competitive performance compared to the popular group lasso, logic regression and Bayesian QTL mapping methods in the literature.