Selective inference after feature selection via multiscale bootstrap

Yoshikazu Terada[1] · Hidetoshi Shimodaira[2,3]

Abstract

It is common to show the confidence intervals or *p*-values of selected features, or predictor variables in regression, but they often involve selection bias. The selective inference approach solves this bias by conditioning on the selection event. Most existing studies of selective inference consider a specific algorithm, such as Lasso, for feature selection, and thus they have difficulties in handling more complicated algorithms. Moreover, existing studies often consider unnecessarily restrictive events, leading to over-conditioning and lower statistical power. Our novel and widely applicable resampling method via multiscale bootstrap addresses these issues to compute an approximately unbiased selective *p*-value for the selected features. As a simplification of the proposed method, we also develop a simpler method via the classical bootstrap. We prove that the *p*-value computed by our multiscale bootstrap method is more accurate than the classical bootstrap method. Furthermore, numerical experiments demonstrate that our algorithm works well even for more complicated feature selection methods such as non-convex regularization.

✉ Yoshikazu Terada
 terada@sigmath.es.osaka-u.ac.jp

 Hidetoshi Shimodaira
 shimo@i.kyoto-u.ac.jp

1 Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531, Japan

2 Graduate School of Informatics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

3 Jointly affiliated at RIKEN Center for Advanced Intelligence Project (AIP), 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan