A Model-averaging method for high-dimensional regression with missing responses at random

Niansheng Tang

Yunnan University
E-mail: nstang@ynu.edu.cn

Abstract: This article considers the ultrahigh-dimensional prediction problem in the presence of missing responses at random. A two-step model averaging procedure is proposed to improve prediction accuracy of conditional mean of response variable. The first step is to specify several candidate models, each with low-dimensional predictors. To implement this step, a new feature screening method is developed to distinguish from the active and inactive predictors via the inverse probability weighted rank correlation (IPWRC), and candidate models are formed by grouping covariates with similar size of IPWRC values. The second step is to develop a new criterion to find the optimal weights for averaging a set of candidate models via the weighted delete-one cross-validation (WDCV). Under some regularity conditions, we show that the proposed new screening statistic enjoys ranking consistency property, and the WDCV criterion asymptotically achieves the lowest possible prediction loss. Simulation studies and an example are illustrated by the proposed methodologies.