Sparse SIR via Lasso

Qian lin

Tsinghua University
E-mail: qianlin@mail.tsinghua.edu.cn

Abstract: For multiple index models, it has recently been shown that the sliced inverse regression (SIR) is consistent for estimating the sufficient dimension reduction (SDR) space if and only if $\rho=\lim pn=0$, where p is the dimension and n is the sample size. Thus, when p is of the same or a higher order of n, additional assumptions such as sparsity must be imposed in order to ensure consistency for SIR. By constructing artificial response variables made up from top eigenvectors of the estimated conditional covariance matrix, we introduce a simple Lasso regression method to obtain an estimate of the SDR space. The resulting algorithm, Lasso-SIR, is shown to be consistent and achieve the optimal convergence rate under certain sparsity conditions when p is of order $o(n2\lambda 2)$, where λ is the generalized signal-to-noise ratio. We also demonstrate the superior performance of Lasso-SIR compared with existing approaches via extensive numerical studies and several real data examples.