
S. EJAZ AHMED, Brock U

Implicit BIAS Problems in in High Dimensional Predictive Models: "Cut the Bias"

The rapid growth in the size and scope of data sets in a variety of disciplines have naturally led to the usage of the term, Big Data. The analysis of such data is important in multiple research fields such as digital marketing, gene expression arrays, social network modeling, clinical, genetics and phenotypic data, bioinformatics, personalized medicine, environmental, neuroscience, astronomy, nanoscience, among others. In high-dimensional models where number of predicting variables is greater than observations, many penalized regularization strategies were studied for simultaneous submodel selection and post-estimation. Generally speaking, submodel are subject to inherited bias, and the prediction based on a selected submodel may not be preferable. For this reason, we propose a high-dimensional shrinkage strategy to improve the prediction performance of a submodel. Such a high-dimensional shrinkage estimator (HDSE) is constructed by shrinking a overfitted model estimator in the direction of a candidate submodel. We demonstrate that the proposed HDSE performs uniformly better than the overfitted model estimator. Interestingly, it improves the prediction performance of a given candidate submodel. The relative performance of the proposed HDSE strategy is appraised by both simulation studies and the real data analysis.