

Multiple influential point detection in high dimensional regression spaces

Junlong Zhao

School of Statistics, Beijing Normal University
E-mail: zhaojunlong928@126.com

Abstract: Influence diagnosis is an integrated component of data analysis but has been severely underinvestigated in a high dimensional regression setting. One of the key challenges, even in a fixed dimensional setting, is how to deal with multiple influential points that give rise to masking and swamping effects. The paper proposes a novel group deletion procedure referred to as multiple influential point detection by studying two extreme statistics based on a marginal-correlation-based influence measure. Named the min- and max-statistics, they have complementary properties in that the max-statistic is effective for overcoming the masking effect whereas the min-statistic is useful for overcoming the swamping effect. Combining their strengths, we further propose an efficient algorithm that can detect influential points with a prespecified false discovery rate. The influential point detection procedure proposed is simple to implement and efficient to run and enjoys attractive theoretical properties. Its effectiveness is verified empirically via extensive simulation study and data analysis. An R package implementing the procedure is freely available.