Data-driven selection of the number of jumps in regression curves: consistency and error rate control

Guanghui Wang

Nankai University
E-mail: ghwang.nk@gmail.com

Abstract: In nonparametric regression with jump discontinuities, one of the major challenges is to determine the number of jumps. Most existing approaches are based on sequential tests or are derived from the model selection viewpoint, which inevitably introduce additional tuning parameters that may not be robust for practical use. We develop a data-adaptive framework with the help of an order-preserved sample-splitting strategy. A cross-validation-based criterion is proposed and its selection consistency is established. More interestingly, the proposed framework allows us to move beyond the point estimation---a new selection procedure with uncertainty quantification is proposed. The key idea is to construct a series of statistics with marginal symmetry property and then to utilize the symmetry for constructing a data-driven threshold to control the false discovery rate. The proposed methodology is computationally efficient, and numerical experiments indicate that it is able to deliver more robust detection results than existing methods in finite samples.