A CLUSTERING APPROACH TO BROKEN-LINE REGRESSIONS

[1]KANG-PING LU, [2]SHAO-TUNG CHANG

[1]National Taichung University of Science and Technology, Taiwan,
[2]National Taiwan Normal University, Taiwan,
E-mail: [1]kplu@nutc.edu.tw/kplu5208@gmail.com, [2]schang@math.ntnu.edu.tw/schang4809@gmail.com

Abstract - Broken-line regressions have been commonly considered in real applications in which break-point detection is crucial. In this study, we consider the break-points as the partitioning between two adjoining segments. Accordingly, the clustering technique can be employed for locating break-points. Moreover, many existing methods for broken-line regressions are based on the assumption that the regression errors have normal distributions. In reality, many data setsare in the presence of longer-than-normal tails or atypical observations, the use of normal errors may unduly affect the fit of the broken-line regression model. tregression errors have been considered for robust statistical modeling, but they were rarely used for broken-line regression problems. This study proposes a broken-line regression model with t-distributed errrors.We simplify the computation of the t likelihood function using the property of a t-distributionas a mixture of normal distributions. Implementing the expectation and maximization algorithm, we create an EM-based algorithm to obtain resistant estimates of the break-points and model parameters simultaneously. Simulation studies show the proposed algorithm is resistant to atypical observations and free of the initialization problem. Extensive experiments demonstrate the preference of the t-based method over the normal-based approaches in the view of robustness and computational efficiency. For Gaussian data without noises, the new approach performs as well as the normal-based method. Real applications reveal that the practicability of the proposed method.

Keywords - Broken-Line Regressions, Atypical Observations, Mixture Models, Break-Points