Feature selection for high-dimensional heteroscedastic regression and its application to root cause analysis with manufacturing data

Po-Hsiang Peng[1], Hai-Tang Chiou[2*], Hsueh-Han Huang[1], Ching-Kang Ing[1]

[1]*Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan*
[2]*Department of Mathematics, National Chung Cheng University, Chiayi, Taiwan*

Abstract

We consider feature selection for high-dimensional linear heteroscedastic models. Inspired by the connection between the linear heteroscedastic function and the interaction model, we propose a two-stage algorithm based on the multi-step orthogonal greedy algorithm and the high-dimensional information criterion to choose the relevant features in the aforementioned high-dimensional model. We prove selection consistency of the proposed method and illustrate its performance via numerical simulations and real data analysis.

Keywords: High-dimensional information criterion; High-dimensional interaction model; Orthogonal greedy algorithm.