

Sliced Inverse Regression for the Interval-valued Symbolic Data

Han-Ming Wu

Tamkang University, New Taipei City, Taiwan hmwu@mail.tku.edu.tw

Sliced inverse regression (SIR) was introduced by Li (1991) to find the effective dimension reduction directions for exploring the intrinsic structure of high-dimensional data. For univariate response regression, SIR has been extended and applied to different data types. Examples were the cases of the survival data, the time series data, the functional data and the longitudinal data. This study intends to develop SIR for the interval-valued symbolic data. Firstly, the interval-valued data was transformed into the conventional data matrix using the vertices method or the centers method. Then the classical SIR algorithm was directly applied to the transformed data. The simulation results show that using different slicing schemes produced different projection directions and different lower-dimensional visualization. Therefore, a suitable slicing scheme is needed for correctly investigating the embedded structure and information of the high-dimensional interval-valued symbolic data in the lower-dimensional plots. The results motivated us to adopt the clustered-based SIR to improve the implementation of the symbolic SIR. We compared and evaluated the results with those obtained with several existing symbolic dimension reduction techniques (such as the symbolic principal component analysis) for discriminative and visualization purposes.

Key Words: Data visualization, sufficient dimension reduction, symbolic principal component analysis.