Nonlinear Dimension Reduction Through Cumulative Slicing Estimation for Nonlinear Manifolds Learning

Han-Ming Wu
hmwu@mail.tku.edu.tw
Department of Mathematics, Tamkang University, Taiwan

Sliced inverse regression (SIR) was developed to find the effective dimension reduction directions for exploring the intrinsic structure of high-dimensional data. The isometric SIR (ISOSIR), a nonlinear extension of SIR, employed K-means on the pre-calculated isometric distance matrix of the data set so that the classical SIR algorithm can be applied. It has been shown that ISOSIR can recover the embedded dimensionality and the geometric structure of the nonlinear manifolds data sets such as the Swiss roll. However, by using K-means, ISOSIR ignored the ordering information of response $y$s both within and between the slices where the ordering information was one of the most important characteristics of the nonlinear manifold data sets. In this study, we are motivated to settle this problem by using the cumulative slicing estimation. First, the proposed method computes the isometric distance between data points; the resulting distance matrix is then sorted by the rank-two ellipse seriation method, and the classical cumulative slicing estimation algorithm is applied. We conducted a pilot study and shown that the proposed method can reveal the geometric structure of a nonlinear manifold data set such as the Swiss roll and the results were comparable to ISOSIR. We then discussed the statistical properties of the proposed method and address how to obtain the response as an estimation of the ordering structure of the data when it was not available. We also investigated the applications of the found features for the classification, clustering and regression problems to the real world data and microarray gene expression data. The comparisons with those obtained with several existing nonlinear dimension reduction techniques were also examined.