Dimension Reduction and Visualization of Histogram Data Using Sliced Inverse Regression

Han-Ming Wu
Department of Mathematics, Tamkang University

Abstract

The dimension reduction of the histogram-valued data is one of the active research topics in symbolic data analysis (SDA). The main thread has been focused on the extensions of the principal component analysis (PCA) though. In this study, we extended the classical sliced inverse regression (SIR), an alternative method to dimension reduction, to the histogram data. SIR was one of the popular sliced-based sufficient dimension reduction techniques for exploring the intrinsic structure of high-dimensional data. We first considered the empirical (joint) density of histogram-valued variables to compute the symbolic weighted covariance-variance matrix. Then a linear combination of histograms rule and the matrix visualization technique were employed to visualize the projections of histograms in the low-dimensional subspace. We evaluated the method for the low-dimensional discriminative and visualization purposes by an application to several real datasets. The comparison with those obtained with the histogram PCA was also reported.

Keywords: Data visualization, histogram-valued data, sliced inverse regression, sufficient dimension reduction, symbolic principal component analysis.