

Principal Subsets Analysis

Maryam Tayefi and Sharad Gore
Department of Statistics
University of Pune
Pune 411 007, India

Abstract

Multivariate data are always difficult to handle due to the so-called curse of dimensionality. Researchers have developed method for reducing the dimensionality of multivariate data with help of mathematical transformations. Principal component analysis, factor analysis, and independent component analysis are examples of such methods that achieve a reduction in dimension of data elements. One of the major drawbacks of these methods is that the resulting dimensions are difficult to interpret because they are mathematical constructs and are not observed characteristics of the population units.

A new method for reducing the dimension by forming groups of highly correlated dimensions is introduced in this paper. It uses the correlation coefficients as the measure of association between the dimensions of data elements and forms a partition of these dimensions. Since there is no mathematical transformation, the dimensions continue to be as observed and hence the results are interpretable. How to use the partition for further processing of the data is a question addresses in the paper. Examples are given to explain and illustrate the new method, called the Principal Subsets Analysis (PSA).