Fuzzy Subspace Skeletons and Applications to Blind Signal Separation

PANDO GEORGIEV, ANCA RALESCU
Electrical and Computer Engineering and Computer Science Department
University of Cincinnati, Cincinnati, OH 45220
Email: pgeorgie@ececs.uc.edu, aralescu@ececs.uc.edu

DAN RALESCU
Department of Mathematical Sciences
University of Cincinnati, Cincinnati, OH 45220
Email: ralescd@ececs.uc.edu

Abstract

We define a fuzzy subspace skeleton of data points and propose an algorithm for finding it. Such a skeleton is connected with data representation: if the data points (represented as columns of a given matrix $\mathbf{X}$) belong exactly to this fuzzy skeleton, then under some mild conditions we can represent $\mathbf{X}$ of the form $\mathbf{X} = \mathbf{AS}$ uniquely (up to scaling and permutation), where the matrices $\mathbf{A}$ and $\mathbf{S}$ with dimensions $m \times n$ and $n \times N$ respectively, $n \geqslant m$ (often called mixing matrix or *dictionary* and source matrix) are such that $\mathbf{S}$ is sparse in sense that each column of $\mathbf{S}$ has at least $n - m + 1$ zero elements. We develop a fuzzy algorithm for clustering over subspaces, which is essential for identification of the mixing matrix $\mathbf{A}$. The idea of this clustering is the same as in the classical fuzzy clustering problem, but instead of balls, here we cluster over subspaces. The dimensionality of each of these subspaces is determined from the eigen-values of a special combined covariance matrix up to some threshold, like in Principal Component Analysis (PCA). When the cluster is one, this procedure can be considered as a Fuzzy PCA. For more clusters, it is a combination of two ideas, resembling the two steps in the classical k-means clustering algorithm: cluster update, which in our realization it is a Fuzzy PCA, and cluster assignment. Even for the case of PCA (instead of fuzzy PCA) the algorithm is new. Application to data representation of genes and other problems from bioinformatics are considered.