Abstract

In many areas of machine learning, computer vision, image processing, data mining, and information retrieval one is confronted with intrinsically low-dimensional data lying on multiple manifolds embedded in a very high dimensional space. Two fundamental tasks associated with modeling high-dimensional data lying on multiple low-dimensional manifolds are dimensionality reduction and clustering.

The goal of this research is to develop a mathematical framework for simultaneous dimensionality reduction and manifold clustering that addresses some of the state-of-the-art challenges and to explore different applications in computer vision and image processing. The proposed framework relies on recent advances in the sparse representation theory, which addresses the problem of recovering a sparse representation of signals in an appropriate basis from a limited number of measurements. State-of-the-art sparse recovery methods address the problem of sparse representation of signals in a single subspace, or in multiple subspaces with a known basis for each subspace. Our goal is to extend these results in several dimensions in order to address the problem of manifold clustering and dimensionality reduction. First, we plan to find efficient methods for finding the sparse representation of data lying on multiple subspaces without knowing a basis for each subspace. Second, we plan to study conditions on the subspace arrangements and the distribution of the data across subspaces under which our proposed sparse recovery method is guaranteed to recover the true sparse representation of the data points. Third, we plan to address the general case of dimensionality reduction and clustering nonlinear manifolds using sparse recovery methods. We evaluate the efficacy of the proposed algorithm for several applications in computer vision and image processing.