

A novel sparse coding algorithm for classification of tumors based on gene expression data

Morteza Kolali Khormuji

Islamic Azad University, Iran

High-dimensional genomic and proteomic data play an important role in many applications in medicine such as prognosis of diseases, diagnosis, prevention and molecular biology, to name a few. Classifying such data is a challenging task due to the various issues such as curse of dimensionality, noise and redundancy. Recently, some researchers have used the sparse representation (SR) techniques to analyze high-dimensional biological data in various applications in classification of cancer patients based on gene expression datasets. A common problem with all SR-based biological data classification methods is that they cannot utilize the topological (geometrical) structure of data. More precisely, these methods transfer the data into sparse feature space without preserving the local structure of data points. In this paper, we proposed a novel SR-based cancer classification algorithm based on gene expression data that takes into account the geometrical information of all data. Precisely speaking, we incorporate the local linear embedding algorithm into the sparse coding framework, by which we can preserve the geometrical structure of all data. For performance comparison, we applied our algorithm on six tumor gene expression datasets, by which we demonstrate that the proposed method achieves higher classification accuracy than state-of-the-art SR-based tumor classification algorithms.

kolalimorteza@yahoo.com