Abstract

The task to capture and interpret information hidden inside high-dimensional data can be considered very complicated and challenging. Usually, dimension reduction technique may be considered as the first step to data analysis and exploration. The focus of this paper is on high-dimensional data dimension reduction using a supervised artificial neural networks technique known as Auto-Associative Neural Networks (AANN). The AANN can be considered as a powerful tool in data analysis and clustering with the ability to deal with linear and nonlinear correlation among variables. This technique is sometimes referred to as nonlinear principal component analysis (NLPCA), Encoding-Decoding networks, or bottleneck neural networks (BNN) due to its unique structure. It reduces high-dimensional data into low-dimensional data on its bottleneck layer which can later be used for data transmission, clustering and visualization. In this paper, a structurally flexible AANN is developed by using high level computer language, applied and studied on two case studies of Iris flowers and Italian olive oils datasets. The purpose of the work was to investigate the ability of AANN to reduce dimension of high-dimensional data on small (Iris) and large (Olive) datasets. The results have shown that AANN has been able to compress high-dimensional data into only one or two non-linear
Dimension Reduction and Clustering of High Dimensional Data using Auto-Associative Neural Networks

principal components at its bottleneck layer with the highest accuracy of 98. 9% and 82. 1% for both datasets respectively. AANN has also managed to perform accurately in both reducing dimension and clustering data by only using small portion of training dataset.

References


**Index Terms**

Computer Science  
Artificial Intelligence

**Keywords**

Dimension Reduction  
Auto-Associative Neural Networks