

Convex and Semi-Nonnegative Matrix Factorizations for Data Clustering

Professor Chris DING

Lawrence Berkeley National Laboratory

Abstract:

Matrix factorization based approach such as principal component analysis/singular value decompositions is going through a renaissance because of their usefulness in unsupervised learning and low-dimensional approximations. (Many prominent mathematicians, such as Steve Smale and David Mumford, are working on these challenging problems).

Nonnegative matrix factorization (NMF) factorizes a nonnegative input matrix X into two nonnegative matrix factors $X = FG$. Although it is initially proposed for "parts-of-whole" decomposition, we show that NMF is equivalent to the well-known K-means clustering: G is cluster indicator and F contains cluster centroids. This can be generalized to semi-NMF where X, F contain mixed-sign data, while G being nonnegative. We further propose convex-NMF by restricting F be convex combinations of data points, ensuring F to contain meaningful cluster centroids.

An interesting aspect of NMF is deriving algorithms for computing semi-NMF and convex-NMF using optimization theory and prove their convergence using auxiliary functions with a number of inequalities.

We present experiments on face images, newgroups, web log and text data to show the effectiveness of these NMF based clustering.

Based on joint work with Tao Li, Michael Jordan.