Generalized Low Rank Models

M. Udell, C. Horn, R. Zadeh, and S. Boyd

- Final paper

- Julia implementation

- Spark implementation

Principal components analysis (PCA) is a well-known technique for approximating a data set represented by a matrix by a low rank matrix. Here, we extend the idea of PCA to handle arbitrary data sets consisting of numerical, Boolean, categorical, ordinal, and other data types. This framework encompasses many well known techniques in data analysis, such as nonnegative matrix factorization, matrix completion, sparse and robust PCA, k-means, k-SVD, and maximum margin matrix factorization. The method handles heterogeneous data sets, and leads to coherent schemes for compressing, denoising, and imputing missing entries across all data types simultaneously. It also admits a number of interesting interpretations of the low rank factors, which allow clustering of examples or of features. We propose a number of large scale and parallel algorithms for fitting generalized low rank models which allow us to find low rank approximations to large heterogeneous datasets, and provide two codes that implement these algorithms.