Reduced rank approaches to principal component and factor analyses

FENG YUAN

Principal Component Analysis (PCA) and Factor Analysis (FA), as effective information extraction tools commonly used in the multivariate data analysis field, have been transformed into a variety of expressions in recent years, within which their penalized sparsification has been widely studied in order to accomplish simple structures of the estimated parameters. Most of them are either based on the least squares method or maximum likelihood estimation.

In this thesis, a new perspective for handling PCA and FA problems is investigated, which refers to the theories of reduced rank approximation, so that classic matrix-based PCA and FA models are transformed into regularized reduced rank approximation problems. Furthermore, unlike typical penalized sparse methods, it is the common part of FA, i.e., $F\Lambda'$ (or PC' in the PCA case), that is penalized in the proposed methods. Inspired by prior research, the optimal solutions for these kinds of minimization problems are determined by a soft shrinkage thresholding operator imposed on a diagonal matrix whose diagonal elements are singular values from the original dataset. This technique considers the common part as a whole, and both factor score and loading matrices are penalized simultaneously, rather than treating the common loading matrix as an isolated term as in previous methods.

The numerical evidence suggests that this reduced rank strategy is a worthwhile attempt and it does perform well in both high-dimensional (the number of variables is much bigger than the number of observations) and large-sample (the number of observations is larger than the number of variables) situations. In addition, the sparseness and magnitude of the loading matrix can also be regulated with the adjustment of tuning parameters. (Behavioral statistics)

Key words: PCA, FA, γ norm penalty, Reduced rank approximation, Soft thresholding