

Gradient-based sparse principal component analysis with extensions to online learning

Yixuan Qiu

Carnegie Mellon University
E-mail: yixuanq@andrew.cmu.edu

Abstract: Sparse principal component analysis (SPCA) is an important technique for dimensionality reduction of high-dimensional data. However, most existing SPCA algorithms are based on non-convex optimization, which provide little guarantee on the global convergence. SPCA algorithms based on a convex formulation, for example the Fantope projection and selection (FPS) model, overcome this difficulty, but are computationally expensive. In this work we study SPCA based on the convex FPS formulation, and propose a new algorithm that is computationally efficient and applicable to large and high-dimensional data sets. Nonasymptotic and explicit error bounds are derived for both the optimization error and the statistical accuracy, which can be used for testing and inference problems. We also extend our algorithm to online learning problems, where data are obtained in a streaming fashion. The proposed algorithm is applied to high-dimensional genetic data for the detection of functional gene groups.