Statistics > Machine Learning

Sparse PCA through Low-rank Approximations

Dimitris S. Papailiopoulos, Alexandros G. Dimakis, Stavros Korokythakis

(Submitted on 3 Mar 2013)

We introduce a novel algorithm that computes the k-sparse principal component of a positive semidefinite matrix A. Our algorithm is combinatorial and operates by examining a discrete set of special vectors lying in a low-dimensional eigen-subspace of A. We obtain provable approximation guarantees that depend on the spectral profile of the matrix: the faster the eigenvalue decay, the better the quality of our approximation. For example, if the eigenvalues of A follow a power-law decay, we obtain a polynomial-time approximation algorithm for any desired accuracy. We implement our algorithm and test it on multiple artificial and real data sets. Due to a feature elimination step, it is possible to perform sparse PCA on data sets consisting of millions of entries in a few minutes. Our experimental evaluation shows that our scheme is nearly optimal while finding very sparse vectors. We compare to the prior state of the art and show that our scheme matches or outperforms previous algorithms in all tested data sets.

Comments:	23 pages, 6 figures, 4 tables, submitted for publication
Subjects:	**Machine Learning (stat.ML)**; Information Theory (cs.IT); Learning (cs.LG)
Cite as:	**arXiv:1303.0551 [stat.ML]**
	(or **arXiv:1303.0551v1 [stat.ML]** for this version)

Submission history

From: Dimitris S. Papailiopoulos [view email]

[v1] Sun, 3 Mar 2013 19:08:55 GMT (1269kb,D)