

Title:

Practical Procedures for Dimension Reduction in l_1

Author(s):

Ping Li, Trevor J. Hastie, and Kenneth W. Church

Technical Report number (Dept. of Statistics, Stanford Univ.):

2006-4

Date:

June 2006

Abstract:

We show that an analog of the Johnson-Lindestrauss (JL) lemma for dimension reduction in l_1 can be established using linear projections[?] and nonlinear estimators. Previous studies[?, ?, ?, ?] have proved that no JL lemmas exist for l_1 using linear estimators. We develop two nonlinear estimators including a strictly unbiased estimator and an improved estimator based on the maximum likelihood. While the maximum likelihood estimator (MLE) does not have a closed-form density function, we propose highly accurate closed-form approximations.

Sampling is also effective for dimension reduction in l_1 . We apply a sketch-based sampling technique for l_1 dimension reduction, which is a combination of sketching and sampling and is particularly advantageous when the data are sparse.

Our results will be useful for applications concerning pairwise l_1 distances, including distance-based clustering, nearest neighbor searching, as well as approximating l_1 kernels for (e.g.) support vector machines (SVM).