Title:
Estimators and Tail Bounds for Dimension reduction in l_α $(0 < \alpha \leq 2)$ Using Stable Random Projections

Author(s):
Ping Li

Technical Report number (Dept. of Statistics, Stanford Univ.):
2007-1

Date:
April 2007

Abstract:

The method of *stable random projections* is popular in data stream computations, data mining, information retrieval, and machine learning, for efficiently computing the l_α $(0 < \alpha \leq 2)$ properties (norms or distances). This study mainly concerns the estimation task. We recommend various estimators depending α, based on the mean square error criterion. A *harmonic mean* estimator is recommended when $0 < \alpha \leq 0.344$. A *biased geometric mean* estimator is suggested for $0.344 < \alpha \leq 1$ and an *unbiased geometric mean* estimator for $1 < \alpha < 2$. Of course, it is well-known that one should use the *arithmetic mean* estimator when $\alpha = 2$.

The (unbiased) *geometric mean* estimator is convenient for theoretical analysis. We derive its tail bounds in explicit exponential forms, for general $0 < \alpha \leq 2$. Consequently, we establish the Lemma that the sample size $k = O\left(\frac{\log n}{\epsilon^2}\right)$ suffices to ensure any l_α distance among n data points can be approximated within a $1 \pm \epsilon$ factor with high probability.