

December 2009

MADALGO seminar by Morteza Monemizadeh, University of Dortmund, Germany

Coresets and Sketches for High Dimensional Subspace Approximation Problems

Abstract:

We consider the problem of approximating a set P of n points in \mathbb{R}^d by a j -dimensional subspace under the l_p measure, in which we wish to minimize the sum of l_p distances from each point of P to this subspace. More generally, the $F_q(l_p)$ -subspace approximation problem asks for a j -subspace that minimizes the sum of q th powers of l_p -distances to this subspace, up to a multiplicative factor of $(1 + \epsilon)$.

We develop techniques for subspace approximation, regression, and matrix approximation that can be used to deal with massive data sets in high dimensional spaces. In particular, we develop coresets and sketches, i.e. small space representations that approximate the input point set P with respect to the subspace approximation problem. Among the results we propose a dimensionality reduction method for various clustering measures, strong coreset, ptas and streaming algorithms in bounded and unbounded precision model for $F_1(l_2)$ -subspace approximation in high-dimensional spaces.

Joint work with:

Dan Feldman
Christian Sohler
David P. Woodruff