Performance of Johnson-Lindenstrauss Transform for k-Means and k-Medians Clustering

Abstract:

Consider an instance of Euclidean k-means or k-medians clustering. We show that the cost of the optimal solution is preserved up to a factor of $(1+\epsilon)$ under a projection onto a random $O(\log(k/\epsilon)/\epsilon^2)$-dimensional subspace whp. Further, the cost of every clustering is preserved within $(1+\epsilon)$ whp. Crucially, the dimension does not depend on the total number of points $n$ in the instance. Additionally, our result applies to Euclidean k-clustering with the distances raised to the $p$-th power for any constant $p$.

For k-means, our result resolves an open problem posed by Cohen, Elder, Musco, Musco, and Persu (STOC 2015); for k-medians, it answers a question raised by Kannan.

Joint work with Konstantin Makarychev and Ilya Razenshteyn.