Accepted paper with abstract, SoCG'09

Andrea Vattani

k-means Requires Exponentially Many Iterations Even in the Plane

The *k*-means algorithm is a well-known method for partitioning n points that lie in the d-dimensional space into k clusters. Its main features are simplicity and speed in practice. Theoretically, however, the best known upper bound on its running time (i.e. $O(n^{kd})$) is, in general, exponential in the number of points (when $kd = \Omega(n / \log n)$). Recently, Arthur and Vassilvitskii [3] showed a super-polynomial worst-case analysis, improving the best known lower bound from $\Omega(n)$ to $2^{\Omega(\sqrt{n})}$ with a construction in $d = \Omega(\sqrt{n})$ dimensions. In [3] they also conjectured the existence of superpolynomial lower bounds for any $d \geq 2$. Our contribution is twofold: we prove this conjecture and we improve the lower bound, by presenting a simple construction in the plane that leads to the exponential lower bound $2^{\Omega(n)}$.