ROBUST K-MEANS CLUSTERING FOR DISTRIBUTIONS WITH TWO BOUNDED MOMENTS

NIKITA ZHIVOTOVSKY

Abstract: We consider the robust algorithms for the k-means clustering problem where a quantizer is constructed based on N independent observations. Our main results are median of means based non-asymptotic excess distortion bounds that hold under the two bounded moments assumption in a general separable Hilbert space. In particular, our results extend the renowned asymptotic result of Pollard who showed that the existence of two moments is sufficient for strong consistency of an empirically optimal quantizer in $\mathbb{R}^d$. In a special case of clustering in $\mathbb{R}^d$, under two bounded moments, we prove matching non-asymptotic upper and lower bounds on the excess distortion, which depend on the probability mass of the lightest cluster of an optimal quantizer. Our bounds have the sub-Gaussian form, and the proofs are based on the versions of uniform bounds for robust mean estimators.
Based on the joint work with Y. Klochkov and A. Kroshnin (https://arxiv.org/abs/2002.02339 to appear in Annals of Statistics).