December 15

16:20 – 17:40
Tuesday

Colloquium
Faculty of Computer Science,
HSE University   №91

Nikita Zhivotovskiy

Google Research, Zurich

On robust mean estimation and k-means clustering

In this talk we consider the robust algorithms for the k-means clustering problem where a quantizer is constructed based on N independent observations. We start with an overview of the methods of robust statistics. First, we discuss the median-of-means estimator. This simple estimator allows us to evaluate the mean of some heavy-tailed distribution as if this distribution was Gaussian. We discuss some extensions to the multivariate case. In the context of clustering, we present the median of means based non-asymptotic distortion bounds that hold under the two bounded moments assumption. 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 R^d. In a special case of clustering in R^d, under two bounded moments, we show matching non-asymptotic upper and lower bounds on the distortion, which depend on the probability mass of the lightest cluster of an optimal quantizer. This talk is based mainly on the joint work with Y. Klochkov and A. Kroshnin (https://arxiv.org/abs/2002.02339 to appear in Annals of Statistics).

Zoom registration link:
https://zoom.us/webinar/register/WN_TBDmutt1QA23RLyX6yHYUA