K-bMOM: a robust K-means-type procedure with application to color quantization.

Édouard Genetay

Classical clustering methods, such as K-means, suffer from a lack of robustness with respect to outliers. We propose a robust version of K-means named K-bMOM, using bootstrap and median-of-means statistics, a strategy that has been recently put to emphasis for efficient robust machine learning. Our algorithm is iterative, in a Lloyd-type fashion. The performances of K-bMOM are theoretically and empirically shown. First, we give a theoretical majoration of the risk excess. Secondly, simulations show that K-bMOM converges rapidly along the iteration steps, that it clearly outperforms K-means on corrupted or heavy-tailed data and that it is competitive with other robust approaches, such as K-median for instance. K-bMOM also provides interesting outcomes such as a robust and efficient initialisation procedure and an outlier detection. This is a joint work with Adrien Saumard and Camille Saumard.