

A new algorithm for robust affine-invariant clustering with equal covariances

Andrews T. Anum^{*} and Michael Pokojovy[†]

^{*}School of Mathematical and Physical Sciences, University of New England, Biddeford, ME

[†]Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA

Abstract

Multivariate datasets typically contain subgroups that may not be immediately apparent. To reveal these groups, cluster analysis is performed. Despite overwhelming popularity of the k -means algorithm, it is neither invariant under non-singular affine transformations nor robust, i.e., can be unduly influenced by outliers. To address these deficiencies, we propose an alternative clustering procedure based on minimizing a “trimmed” variant of the negative log-likelihood function. We develop a “concentration step” that iteratively reduces the objective function to a local minimum in a finite number of steps. Assuming an equal covariance structure amongst clusters, a combined estimate of the covariance known as the pooled covariance is used in place of the individual cluster covariances. Multiple real and synthetic datasets are analyzed to assess the performance of our algorithm based on several cluster quality metrics. Compared to trimmed k -means and `tclust`, empirical studies indicate competitiveness and oftentimes superiority of our algorithm.

Keywords: Cluster analysis; C-step; affine invariance; robustness; equal covariances.