
MEHRDAD GHADIRI, Georgia Institute of Technology
Socially Fair k-Means Clustering

We show that the popular k -means clustering algorithm (Lloyd's heuristic), used for a variety of scientific data, can result in outcomes that are unfavorable to subgroups of data (e.g., demographic groups). Such biased clusterings can have deleterious implications for human-centric applications such as resource allocation. We present a fair k -means objective and algorithm to choose cluster centers that provide equitable costs for different groups. The algorithm, Fair-Lloyd, is a modification of Lloyd's heuristic for k -means, inheriting its simplicity, efficiency, and stability. In comparison with standard Lloyd's, we find that on benchmark datasets, Fair-Lloyd exhibits unbiased performance by ensuring that all groups have equal costs in the output k -clustering, while incurring a negligible increase in running time, thus making it a viable fair option wherever k -means is currently used.