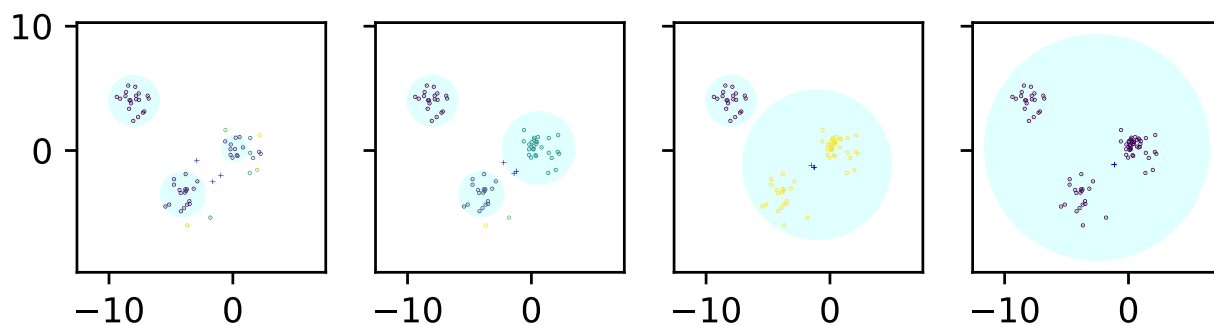


# Convex Clustering: Convexity, Bounding Balls and General Characteristics

Bio-knowledge Engineering Laboratory  
○ Canh Hao Nguyen, Hiroshi Mamitsuka

Convex clustering is an attractive clustering algorithm with favorable properties such as efficiency and optimality owing to its convex formulation. It is thought to generalize both k-means clustering and agglomerative clustering. However, it is not known whether convex clustering preserves desirable properties of these algorithms. A common expectation is that convex clustering may learn difficult cluster types such as non-convex ones. Current understanding of convex clustering is limited to only consistency results on well-separated clusters, which are necessary conditions.

We prove important properties of the solutions of convex clustering (necessary conditions), contrasting them from other clustering algorithms. Contrary to common expectation, we show that convex clustering can only learn convex clusters, unlike agglomerative clustering. We further show that the clusters can be bounded by disjoint bounding balls with radii depending on their sizes. We can say that convex clustering produces circular clusters. Importantly, there are always significant gaps among the bounding balls. This shows a fundamental difference of convex clustering from k-means and other partition-based clustering algorithms that fill up the space. We further show general characteristics on: 1) the samples that result in the same solutions by convex clustering, 2) intuitive guidelines on hyperparameter setting, 3) a case that is impossible to cluster and 4) a guideline to achieve statistical consistency. Figure 1 shows an example of solutions, when more samples are added to the right most component, from left to right.



**Figure 1:** visualization of clustering solutions on datasets with similar cluster structure.