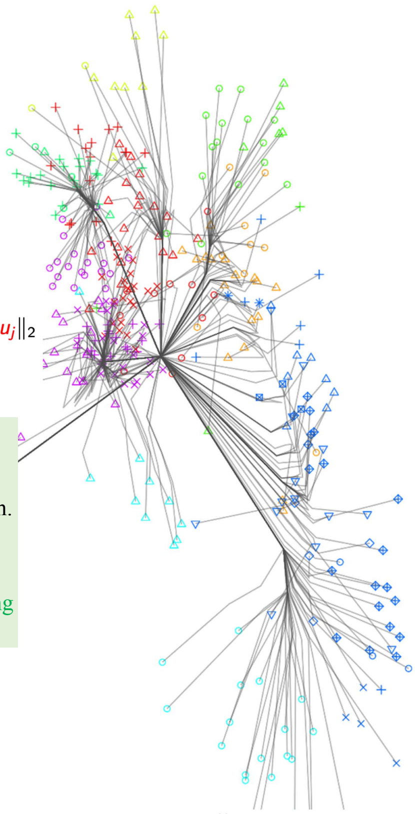
Convex Clustering: Convexity, Bounding Balls and General Characteristics

Canh Hao Nguyen & Hiroshi Mamitsuka
Bio-Knowledge Engineering Research Laboratory
Institute for Chemical Research, Kyoto University



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. In this work, we show sufficient conditions of convex clustering. We prove that convex clustering can only learn convex clusters. We then show that the learnt clusters have disjoint bounding balls with significant gaps. We show the differences to k-means and agglomerative clusterings. Furthermore, we characterize the solutions, regularization hyperparameters, inclusterable cases and consistency.

Convex Clustering Formulation

Given $X = \{x_1, \cdots x_n\} \in \mathcal{R}^{d\times n}$

- Solve
  $\bar{U} = \arg\min_{u_i \in \mathbb{R}^d} \frac{1}{2}\sum_{i=1}^{n} \|x_i - u_i\|_2^2 + \lambda \sum_{j<i} \|u_i - u_j\|_2$
- If $u_i = u_j$ then group x_i, x_j into a cluster.

Formulation: attractive properties

- Generalized k-means clustering:
  - Most commonly used clustering algorithm.
- Generalized agglomerative clustering:
  - Nonconvex cluster solutions.
- Convex formulation: stable, efficient and giving optimal solutions.

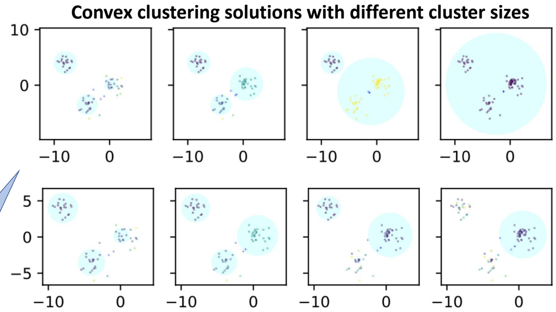
Solutions: good or bad?

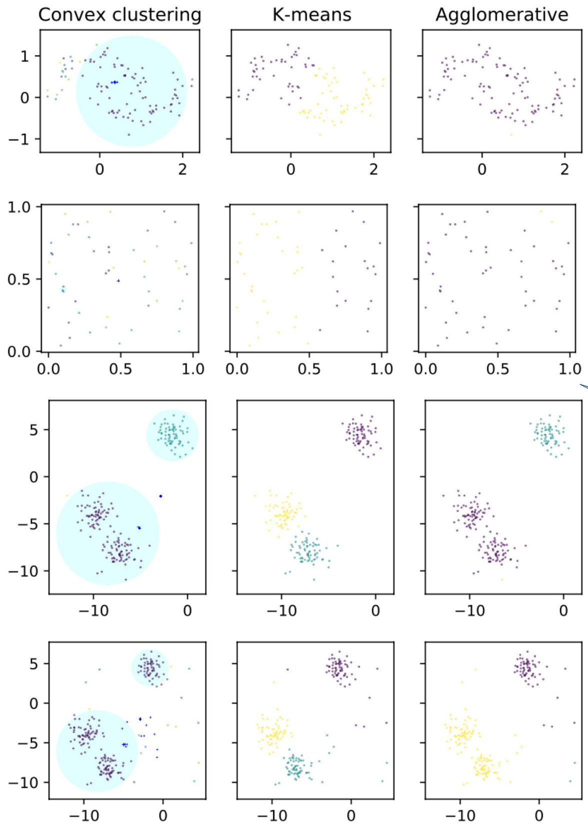
- Can it learn optimal and efficient nonconvex clusters?
- What cluster shapes can it learn?

Main results: we proved that

- Convex clustering can **only learn convex clusters**
  - Different from agglomerative clustering.
  - Similar to k-means clustering.
- **Having bounding balls with significant gaps** (*circular shapes*)
  - Different from k-means clustering (Voronoi cells), no gap.
  - Ball radii proportional to cluster sizes.

Convex clustering solutions with different cluster sizes



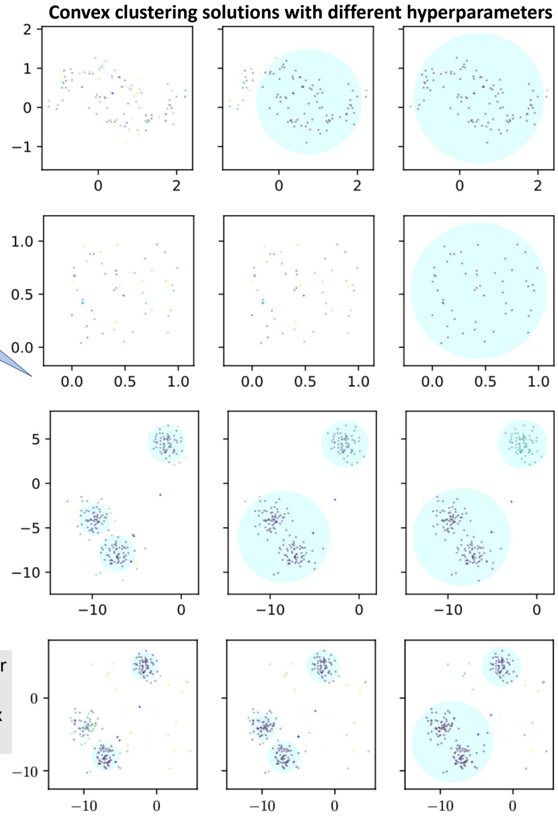


Solutions change by adding new data into the clusters (not robust)

Solutions change quickly with hyperparameters (not stable)

Solutions are different from those of k-means, agglomerative clusterings (unexpected)

Convex clustering solutions with different hyperparameters



- Points in the same cluster have the same color.
- Bounding balls by convex clustering are shaded.

Reference:
Canh Hao Nguyen and Hiroshi Mamitsuka: On Convex Clustering Solutions. *arXiv:2105.08348*