

Skeleton Clustering : A Dimension free Density-Aided Clustering

Kaustav Paul
Sourav Biswas

Indian Statistical Institute, Kolkata

2024-10-29

Traditional Clustering Methods

- **k-means clustering:**

- Unable to detect non-convex clusters.
- The center of a non-convex cluster falls outside the cluster itself and may come close to observations from a different cluster.
- In high dimension k-means algorithm may assign all the points to a single cluster.

- **Density Based Clustering:**

- To estimate the underlying PDF and detect clusters based on the PDF.
- The rate of convergence for the density estimates is $\mathcal{O}_{\mathbb{P}}(n^{-\frac{1}{d+4}})$

- **Hierarchical Clustering:**

- Problem with non-convex clusters persists.
-

Skeleton Clustering Framework.

Input : Observations X_1, X_2, \dots, X_N , final number of clusters S .

- ➊ **Knot construction** : Perform k -means clustering with a large number k ; the centers are the knots.
- ➋ **Edge construction** : Apply approximate Delaunay triangulation to the knots. Generally we choose $k = \lfloor \sqrt{n} \rfloor$
- ➌ **Edge weights construction** : Add weights to each edge using either Voronoi density, Face density or Tube density similarity measure.
- ➍ **Knots segmentation** : Use linkage criterion to segment knots into S groups based on the edge weights.
- ➎ **Assignment of labels** : Assign a cluster label to each observation based on which knot group the nearest knot belongs to.

Knot construction

- Some knots are constructed to give a concise representation of the data structure.
- In practice we use k -Means to choose $k = \lfloor \sqrt{n} \rfloor$ knots, where n is the number of samples.
- Empirically robustness performance with sufficient number of knots.