Skeleton Clustering : A Dimension free Density-Aided Clustering

Kaustav Paul
Sourav Biswas

Indian Statistical Institute, Kolkata

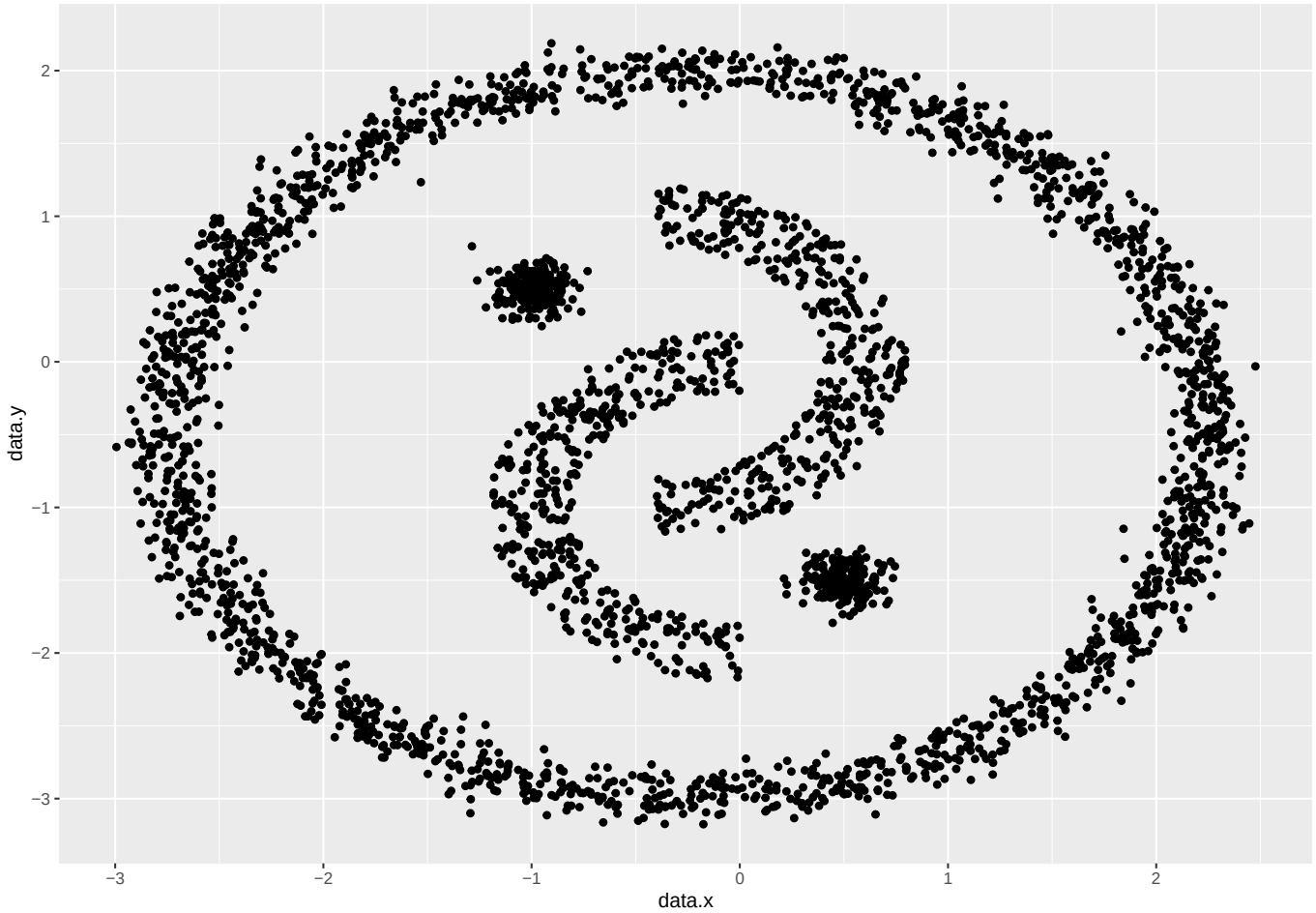
2024-11-05

Traditional Clustering Methods

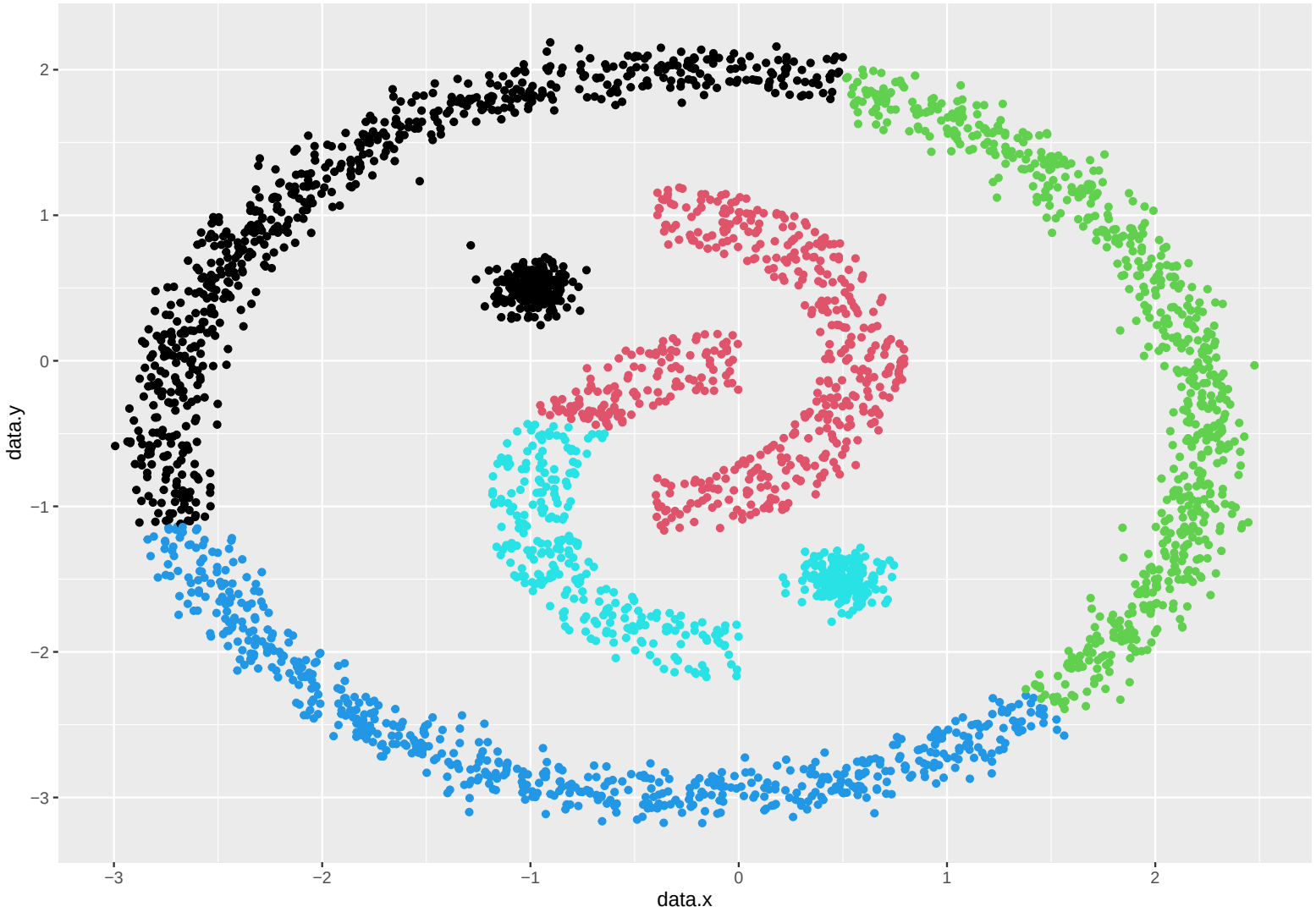
- **k-means clustering:**
  - Unable to detect non-convex clusters.
  - The center of a non-convex cluster falls outside the cluster itself and may come close to observations from a different cluster.
  - In high dimension k-means algorithm may assign all the points to a single cluster.
- **Density Based Clustering:**
  - To estimate the underlying PDF and detect clusters based on the PDF.
  - The rate of convergence for the density estimates is $\mathcal{O}_{\mathbb{P}}(n^{-\frac{4}{d+4}})$
- **Hierarchical Clustering:**
  - In high dimensional problems it may tend to include all points in a single cluster.

The YinYang Data

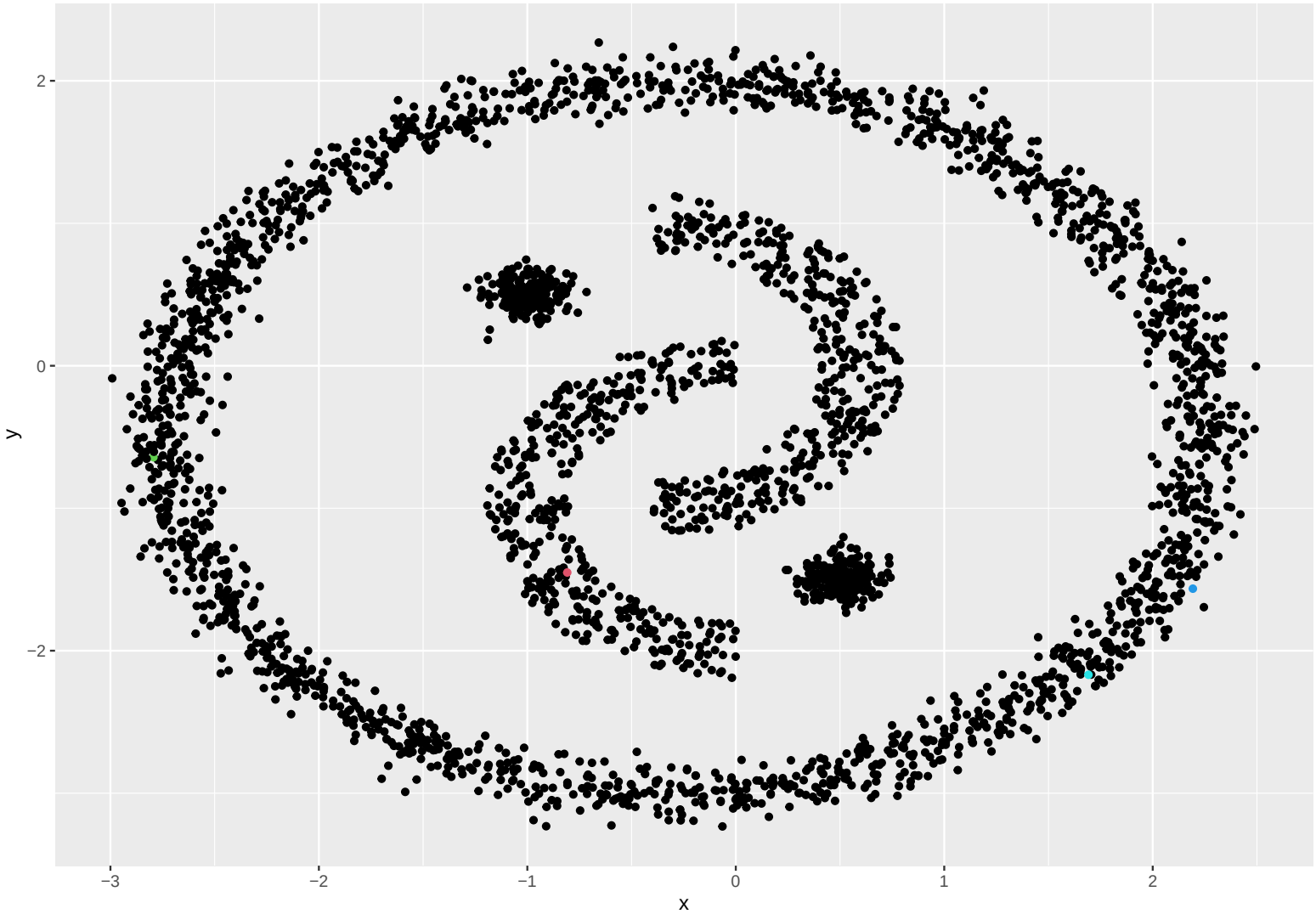
We consider a 100 dimensional data where the first two dimensions have the following structure. The rest 98 are random noise from $N(0, 1)$.



Result of k-means Clustering



Result of Hierarchical Clustering



Skeleton Clustering Framework.

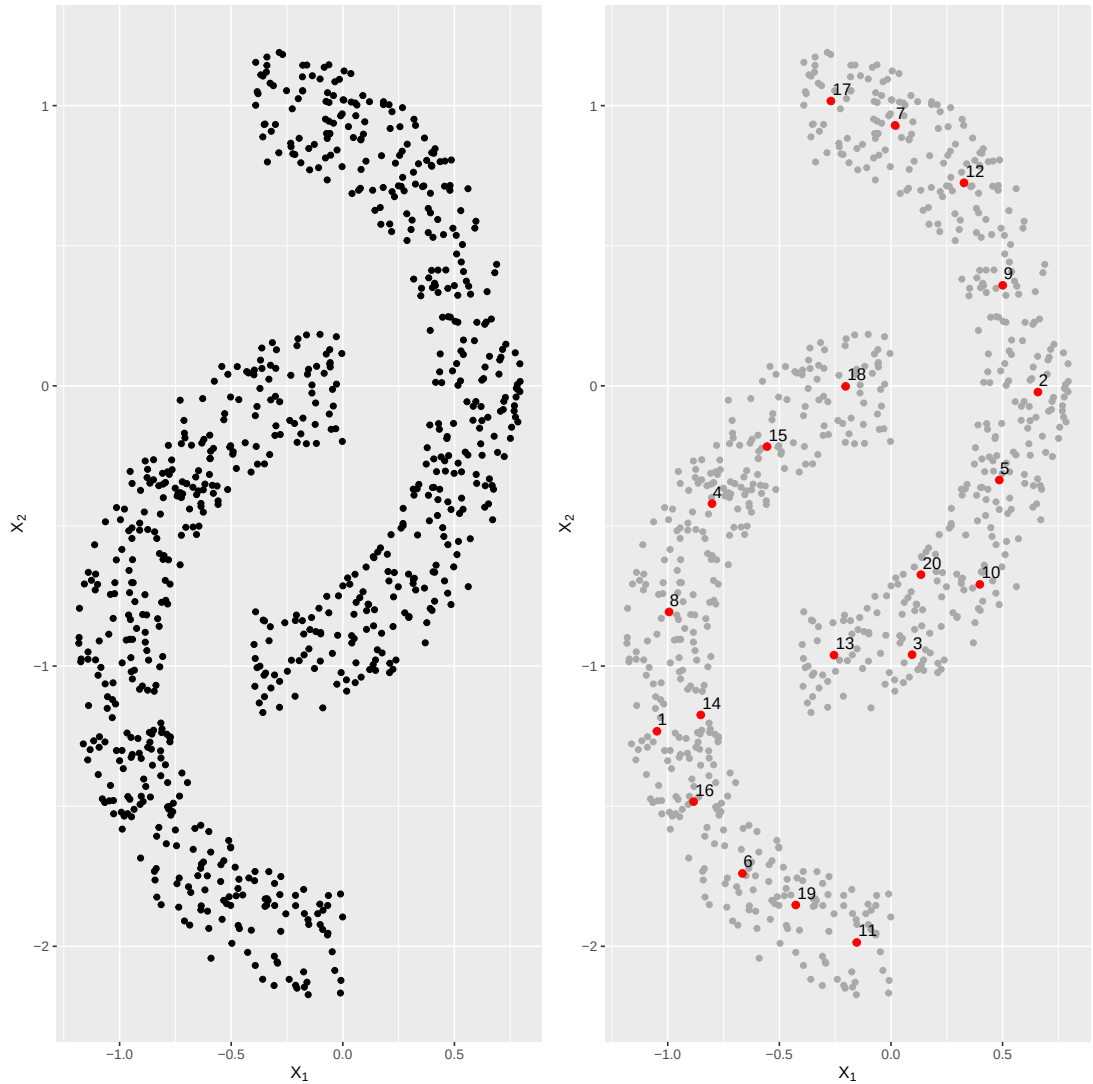
Input : Observations $X_1, X_2, \ldots, X_n$, final number of clusters S.

1. **Knot construction** : Perform $k-$means clustering with a large number k; the centers are the knots.
2. **Edge construction** : Apply approximate Delaunay triangulation to the knots.
3. **Edge weights construction** : Add weights to each edge using either Voronoi density, Face density or Tube density similarity measure.
4. **Knots segmentation** : Use linkage criterion to segment knots into S groups based on the edge weights.
5. **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. One can use k-means algorithm to get such a representation.
- It is suggested to choose $k >>> S$, where S is the final number of clusters. In practice we use $k = \lfloor \sqrt{n} \rfloor$, where n is the number of samples.
- More the number of knots, better will be the representation. However, too many knots will make the density estimation in the later stage problematic.

Knot Construction

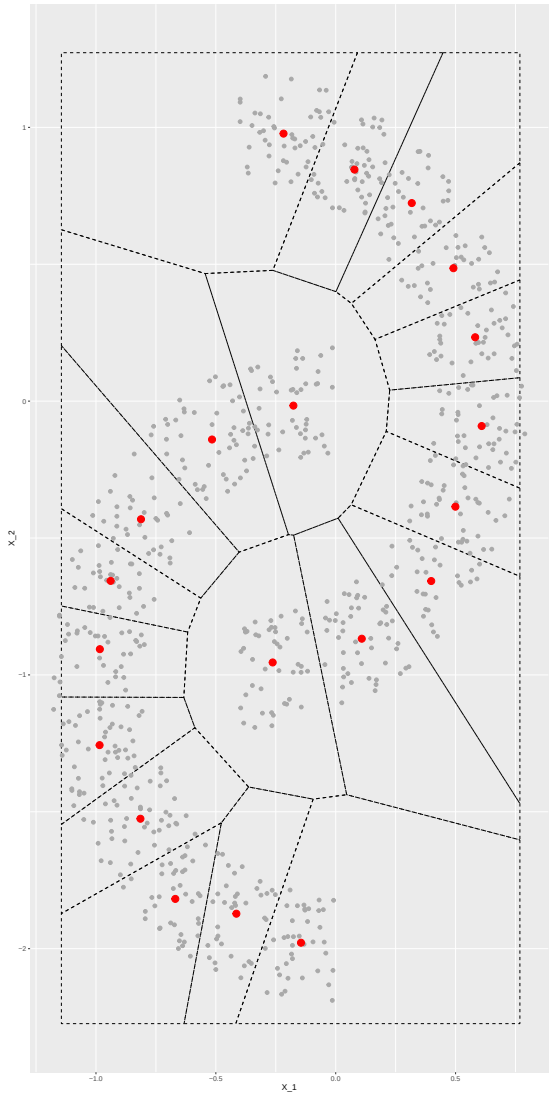


Edge construction

Let $c_1, c_2, \ldots, c_k$ be the given knots and we use $\mathcal{C} = \{c_1, c_2, \ldots, c_k\}$ to denote their collection of them.

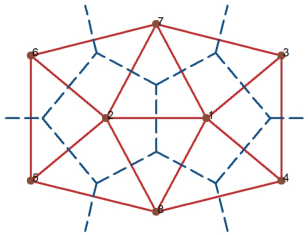
- The **Voronoi cell**, or **Voronoi region**, $\mathbb{C}_j$ associated with a knot c_j is the set of all points in $\mathcal{X}$ whose distance to c_j is the smallest compared to other knots. That is, $\mathbb{C}_j = \{\boldsymbol{x} \in \mathcal{X} : d(\boldsymbol{x}, c_j) \leq d(\boldsymbol{x}, c_\ell) \, \forall \ell \neq j\}$ where $d(\boldsymbol{x}, \boldsymbol{y})$ is the usual Euclidean distance.

Edge Construction



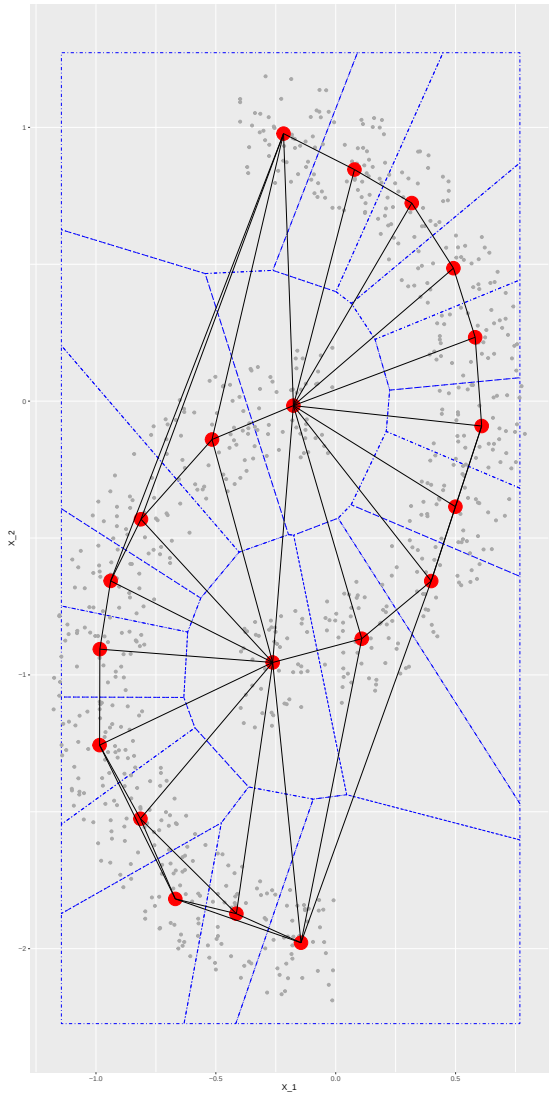
Edge Construction

- We add an edge between a pair of knots if they are neighbors, with the neighboring condition being that the corresponding Voronoi cells share a common boundary.
- Such resulting graph is the **Delaunay Triangulation** of the set of knots $\mathcal{C}$ and we denote is as $DT(\mathcal{C})$.



- But in case of high dimensional data, it becomes computationally expensive. Therefore, in practice we approximate the exact Delaunay Triangulation with $\widehat{DT}(\mathcal{C})$ by examining the 2-nearest knots of the sample data points.

Edge Construction



Edge Weight Construction.

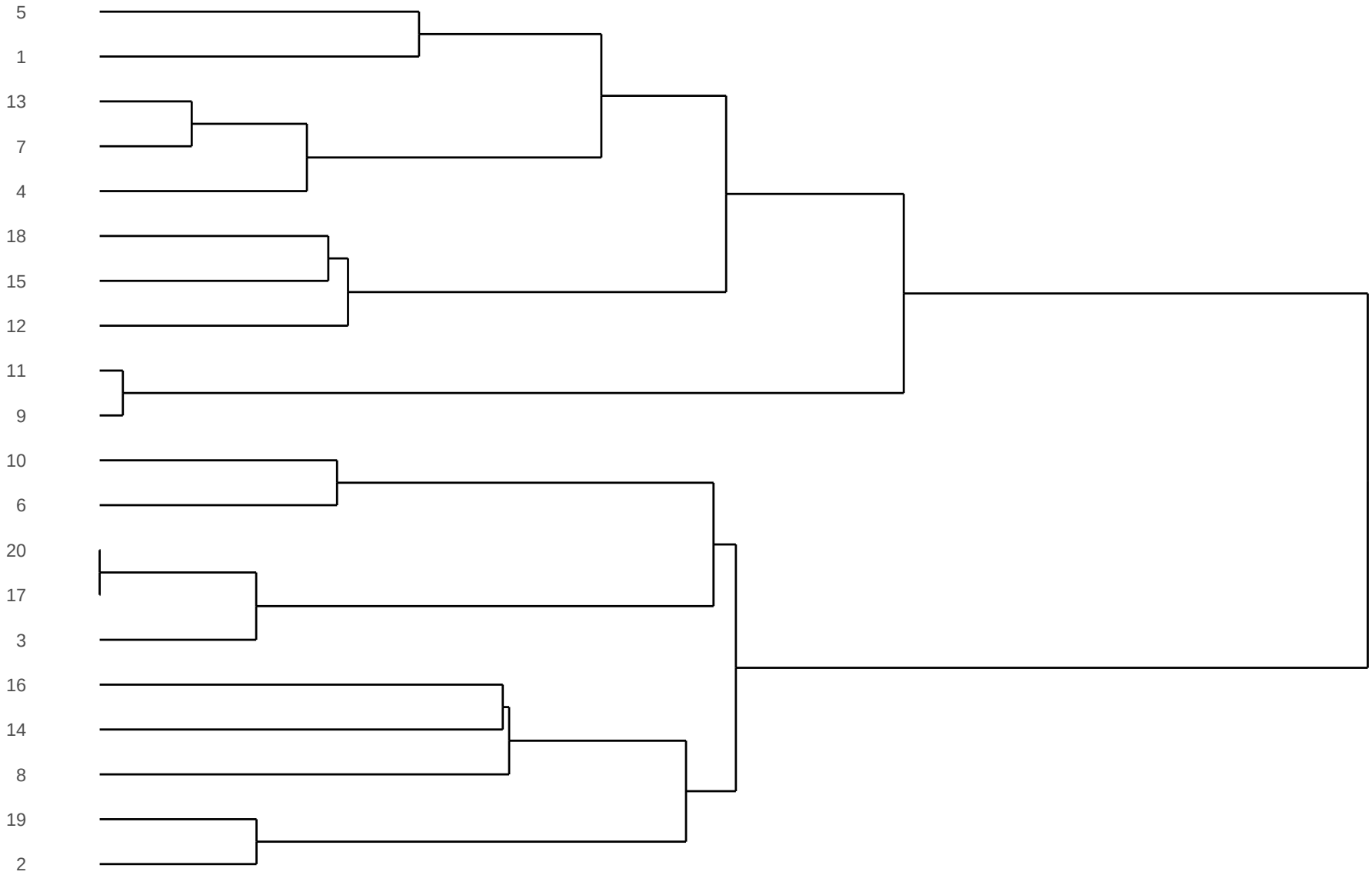
- Given the constructed edges and knots, we assign each edge a weight that represents the similarity between the pair of knots.
- We propose a density aided quantity as edge weights.
- The estimations of the newly proposed similarity measures are reliable even under high-dimensional settings.

Knot Segmentation.

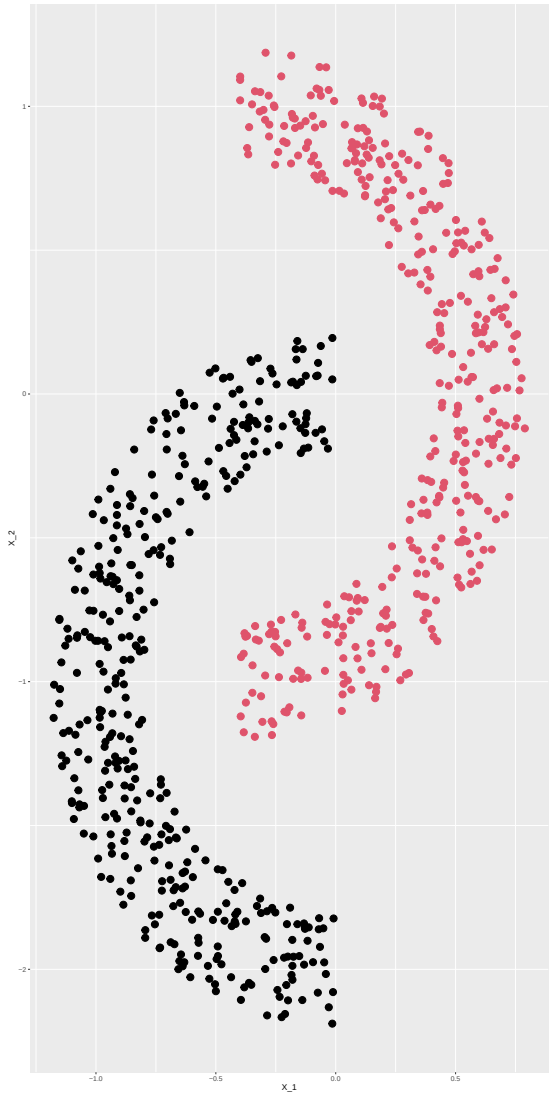
Given the weighted skeleton graph, the next step is to partition the knots into the desired number of final clusters.

- For given similarity measures $\{s_{ij}\}_{i \neq j}$ where only connected pairs can take nonzero entries and let $s_{max} = \max_{i \neq j} s_{ij}$, we define the corresponding distances as $d_{ij} = 0$ if $i = j$ and $d_{ij} = s_{max} - s_{ij}$ otherwise.
- After that, we perform a hierarchical clustering. The choice of the linkage criterion may depend on the underlying geometric structure of the data. In practice, such choice of linkage should be made based on some exploratory understanding of the data structure.

Knot Segmentation



Final Clustering Result



How to choose the final number of Clusters ?

- The number of final clusters S is an essential parameter for the hierarchical clustering procedure but can be unknown.
- An inferential choice can also be made using the gap statistics, Calinski-Harabascz index etc.

Some feasible density functions

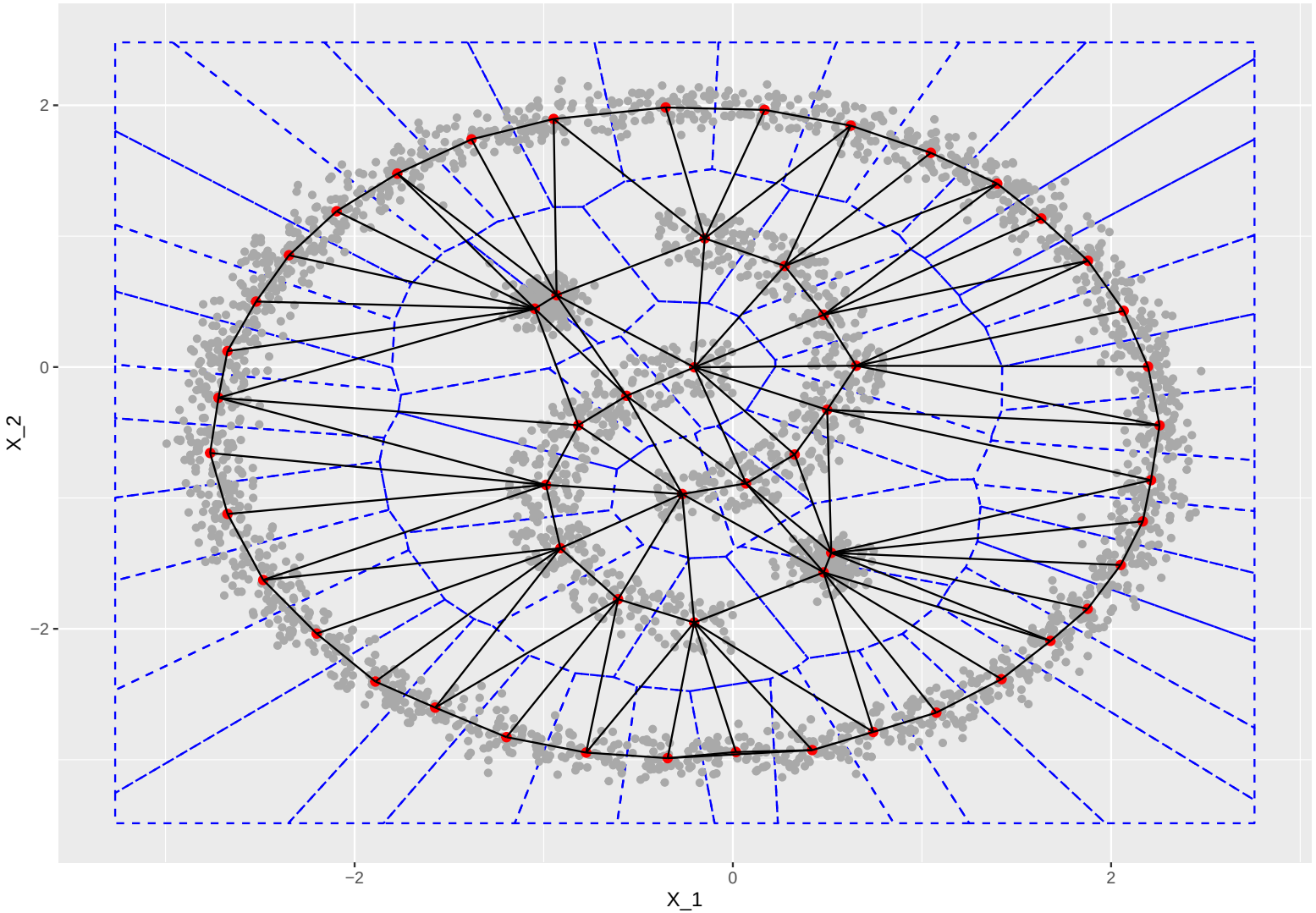
Voronoi Density (VD).

- The Voronoi density (VD) measures the similarity between a pair of knots (c_j, c_ℓ) based on the number of observations whose 2-nearest knots are c_j and c_ℓ.
- Given a metric d on $\mathbb{R}^d$, the 2-Nearest-Neighbor (2-NN) region of a pair of knots (c_j, c_ℓ) is defined as,
  - $$A_{j\ell} = \{\boldsymbol{x} \in \mathcal{X} : d(\boldsymbol{x}, c_i) > \max\{d(\boldsymbol{x}, c_j), d(\boldsymbol{x}, c_\ell)\}\, \forall i \neq j, \ell\}$$
- Following the idea of density-based clustering, two knots c_j, c_ℓ belong to the same clusters if they are in a connected high-density region, and we would expect the 2-NN region of c_j, c_ℓ to have a high probability measure. Hence $\mathbb{P}(A_{j\ell})$ can measure the association between the knots.

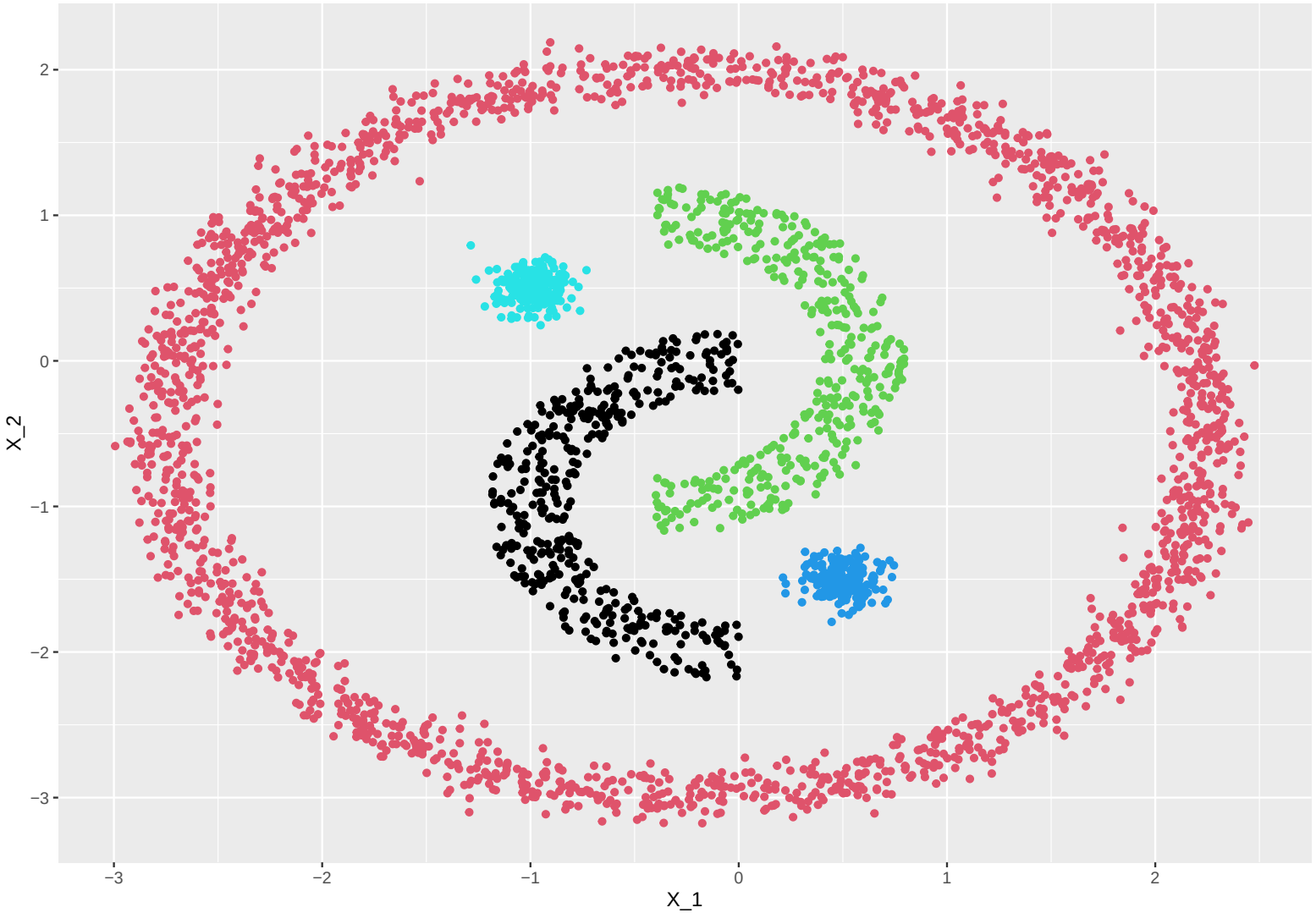
Voronoi Density

- Based on this insight, the Voronoi density measures the edge weight of (c_j, c_ℓ) with,
  - $$S_{j\ell}^{VD} = \frac{\mathbb{P}(A_{j\ell})}{\|c_j - c_\ell\|}$$
- In practice, we estimate $S_{j\ell}^{VD}$ by a sample average. Specifically, the numerator $\mathbb{P}(A_{j\ell})$ is estimated by $\hat{\mathbb{P}}_n(A_{j\ell}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\boldsymbol{X_i} \in A_{j\ell})$ and the final estimator for VD is,
  - $$\hat{S}_{j\ell}^{VD} = \frac{\hat{\mathbb{P}}(A_{j\ell})}{\|c_j - c_\ell\|}$$

Final Result Using Skeleton Clustering



Final Result Using Skeleton Clustering



Asymptotic theory of Edge weight estimation.

We assume the set of knots $\mathcal{C} = \{c_1, \ldots, c_k\}$ is given and non-random to simplify the analysis because

- It is hard to quantify k-means uncertainty
- with large k, it's extremely likely for k-means to stuck within the local minimum.

Voronoi Density consistency

We start with the convergence rate of the VD. We consider the following condition :

- There exists a constant c_0 such that the minimal knot size $\min_{(j,\ell)\in E} \mathbb{P}(A_{j\ell}) \geq \frac{c_0}{k}$ and $\min_{(j,\ell)\in E} \|A_{j\ell}\| \geq \frac{c_0}{k^{\frac{1}{d}}}$

where $(j, \ell) \in E$ means that there is an edge between knots c_j and c_ℓ in the Delaunay Triangulation. The above condition is a condition requiring that no Voronoi cell $A_{j\ell}$ has a particularly small size and all edges have sufficient length.

Voronoi Density Convergence

Theorem:

Under the assumption of the above condition, for any pair $j \neq \ell$ that shares an edge, the similarity measure based on the Voronoi density satisfies,

$$\left| \frac{\hat{S}_{j\ell}^{VD}}{S_{j\ell}^{VD}} - 1 \right| = \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{k}{n}} \right)$$

$$\max_{j,\ell} \left| \frac{\hat{S}_{j\ell}^{VD}}{S_{j\ell}^{VD}} - 1 \right| = \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{k}{n}} \ln k \right)$$

when $n \to \infty, k \to \infty, \frac{n}{k} \to \infty$

References

- Zeyu Wei, Yen-Chi Chen, "Skeleton Clustering : Dimension Free Density-Aided Clustering", 2023.
- J.-P. Baudry, A. E. Raftery, G. Celeux, K. Lo, and R. Gottardo. Combining mixture components for clustering. Journal of Computational and Graphical Statistics, 19(2):332–353, 2010. doi: 10.1198/jcgs.2010.08111.
- A. L. N. Fred and A. K. Jain. Combining multiple clusterings using evidence accumulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6):835–850, 2005. doi: 10.1109/TPAMI.2005.113.
- R. Maitra. Initializing partition-optimization algorithms. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 6(1):144–157, 2009. doi: 10.1109/TCBB.2007.70244.

Thanks for listening! :)