Multi-parameter hierarchical clustering and beyond

Alexander Rolle

Institute of Geometry, TU Graz, Austria

Introduction

- Multi-parameter hierarchical clustering was first introduced by Carlsson–Mémoli [CM10].
- In many applications, the goal is to understand clustering structure across a range of distance scales and density thresholds

Kernel density bifiltration

- Introduced in [RS20], generalizes the degree-Rips bifiltration [LW15].
- Filters a metric probability space according to density estimates computed for all choices of bandwidth parameter.
- Satisfies a stability theorem.

Density-based clustering

- If f is a density function, define the *density-contour hierarchical clustering* by taking connected components of super-level sets.
- Density-based clustering is the problem of approximating this, given finite samples.
- By applying single-linkage to the kernel density bifiltration, and taking one-parameter slices, we obtain an approach to density-based clustering.
- Inherits a stability theorem.

Density estimation

- Density estimation is the problem of approximating the density function f itself.
- There is a related approach to density estimation.
- There is no longer any invariant from topology, like H_0 or π_0, in this application.

Future directions

- Tools from multi-parameter persistence can be used to avoid fixing distance scales or density thresholds.
- Such parameters appear in many geometrically motivated algorithms in data analysis: where else can these tools be applied?

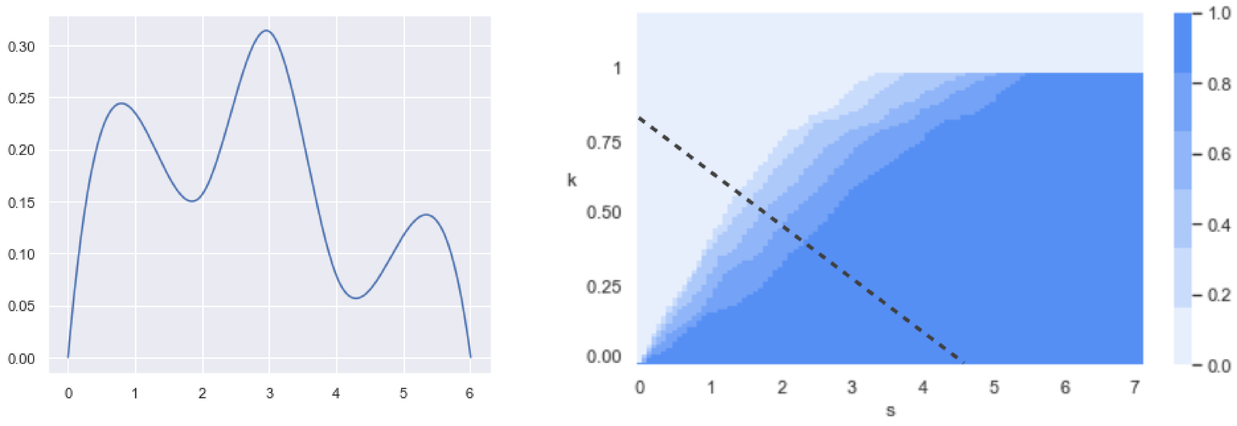

Figure: A density function f on the real line, and a heat map representing the kernel density bifiltration of the support of f. The dashed line is a choice of one-parameter slice.

Acknowledgments

Supported by Austrian Science Fund (FWF) grant P 29984-N35.

References

[CM10] G. Carlsson and F. Mémoli. Multiparameter hierarchical clustering methods. In Hermann Locarek-Junge and Claus Weihs, editors, *Classification as a Tool for Research*, pages 63–70, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[LW15] M. Lesnick and M. Wright. Interactive visualization of 2-D persistence modules. arXiv:1512.00180v1, 2015.

[RS20] A. Rolle and L. Scoccola. Stable and consistent density-based clustering. arXiv:2005.09048, 2020.