

hdbscan: Hierarchical density based clustering

Leland McInnes¹, John Healy¹, and Steve Astels²

¹Tutte Institute for Mathematics and Computing

²Shopify

26 February 2017

Paper DOI: <http://dx.doi.org/10.21105/joss.00205>

Software Repository: <https://github.com/scikit-learn-contrib/hdbscan>

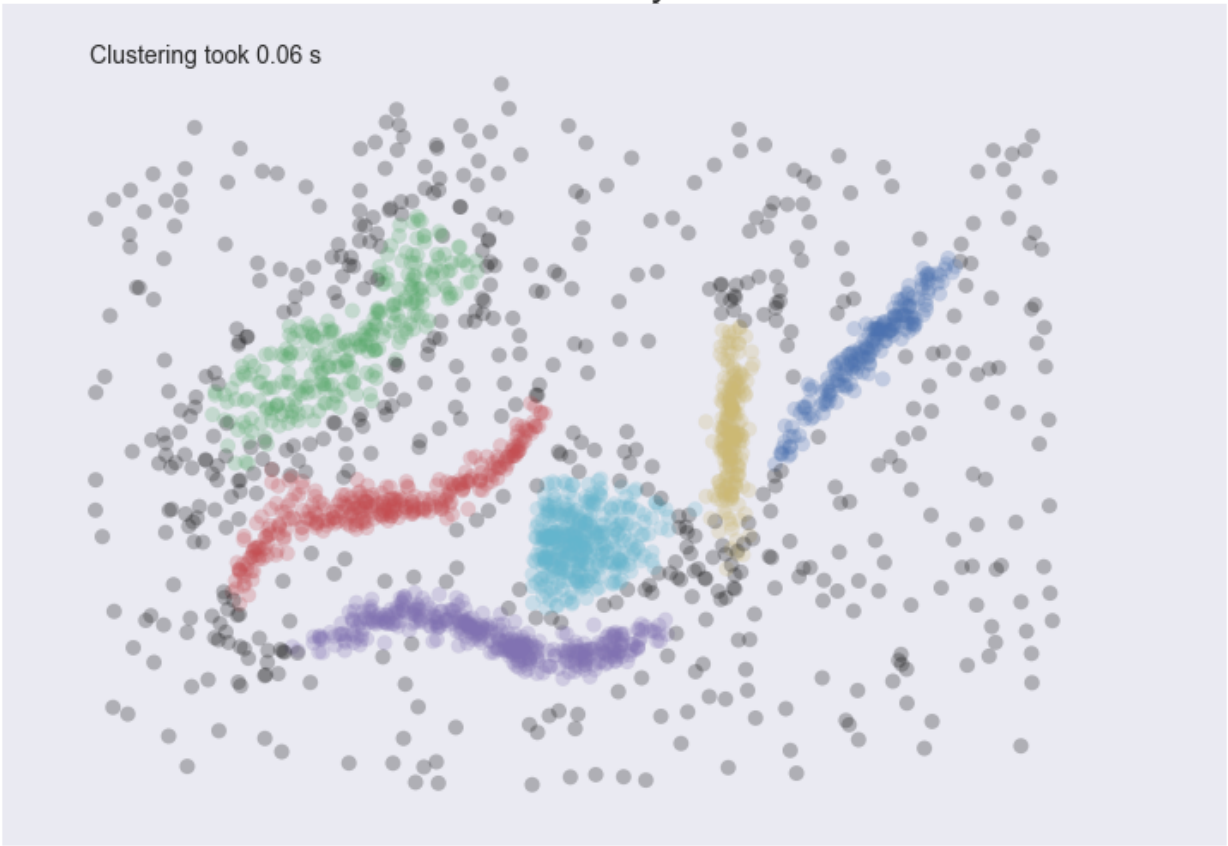
Software Archive: <http://dx.doi.org/10.5281/zenodo.401403>

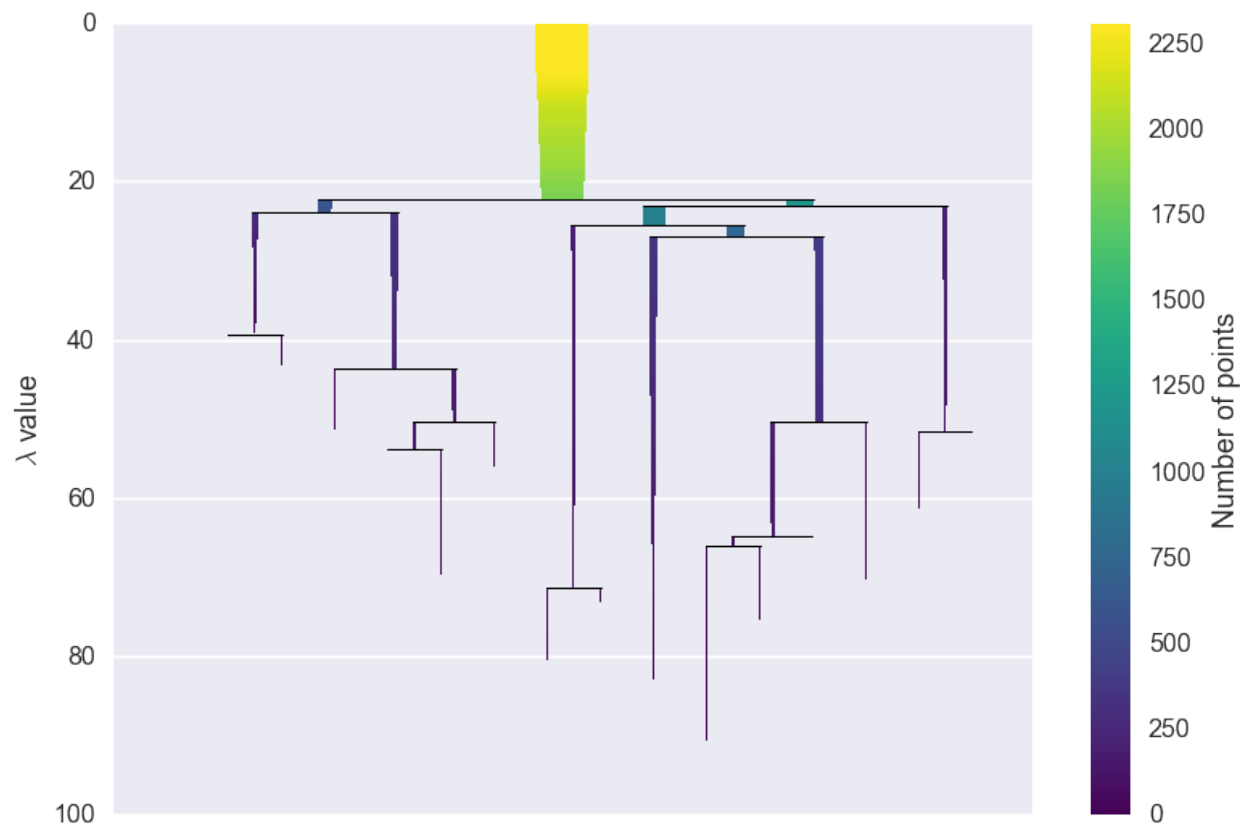
Summary

HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise (Campello, Moulavi, and Sander 2013), (Campello et al. 2015). Performs DBSCAN over varying epsilon values and integrates the result to find a clustering that gives the best stability over epsilon. This allows HDBSCAN to find clusters of varying densities (unlike DBSCAN), and be more robust to parameter selection. The library also includes support for Robust Single Linkage clustering (Chaudhuri et al. 2014), (Chaudhuri and Dasgupta 2010), GLOSH outlier detection (Campello et al. 2015), and tools for visualizing and exploring cluster structures. Finally support for prediction and soft clustering is also available.

Clusters found by HDBSCAN

Clustering took 0.06 s





References

- Campello, Ricardo JGB, Davoud Moulavi, and Joerg Sander. 2013. “Density-Based Clustering Based on Hierarchical Density Estimates.” In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 160–72. Springer. doi:10.1007/978-3-642-37456-2_14.
- Campello, Ricardo JGB, Davoud Moulavi, Arthur Zimek, and Jörg Sander. 2015. “Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection.” *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10 (1). ACM: 5. doi:10.1145/2733381.
- Chaudhuri, Kamalika, and Sanjoy Dasgupta. 2010. “Rates of Convergence for the Cluster Tree.” In *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, 343–51. NIPS’10. USA: Curran Associates Inc. <https://papers.nips.cc/paper/4068-rates-of-convergence-for-the-cluster-tree>.
- Chaudhuri, Kamalika, Sanjoy Dasgupta, Samory Kpotufe, and Ulrike von Luxburg. 2014. “Consistent Procedures for Cluster Tree Estimation and Pruning.” *IEEE Transactions on Information Theory* 60 (12). IEEE: 7900–7912. doi:10.1109/TIT.2014.2361055.