hdbscan: Hierarchical density based clustering

Leland McInnes¹, John Healy¹, and Steve Astels²

¹ Tutte Institute for Mathematics and Computing ² Shopify

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

