mst_clustering: Clustering via Euclidean Minimum Spanning Trees

Jake VanderPlas¹

¹ University of Washington eScience Institute

Summary

This package contains a Python implementation of a clustering algorithm based on an efficiently-constructed approximate Euclidean minimum spanning tree (described in (Ivezić et al. 2014)). The method produces a Hierarchical clustering of input data, and is quite similar to single-linkage Agglomerative clustering. The advantage of this implementation is the ability to find significant clusters even in the presence of background noise, and is particularly useful for researchers hoping to detect structure in physical data.

The code makes use of tools within SciPy (Jones et al. 2001–2001--) and scikit-learn (Pedregosa et al. 2011), and is designed for compatibility with the scikit-learn API (Buitinck et al. 2013).

References


