

A Consistent Density-Based Clustering Algorithm and its Application to Microstructure Image Segmentation

Marilyn Vazquez, ICERM

Data clustering is a fundamental task for discovering patterns in data, and is central to machine learning. Often, a data set is assumed to live in a manifold and be sampled according to a probability measure. Then the clusters can be defined as peaks in the sampled probability density, and a clustering algorithm would need to identify the peaks in the density to compute the clusters. Some of the challenges in this approach include the non-uniform sampling of the density and the bridges between peaks of the density. To solve these problems, we propose a new clustering algorithm that divides the clustering problem into three steps: picking a good threshold on the sample density to separate the peaks, clustering the superlevel set at the chosen threshold, and classifying the remaining points. We explain the key details of these steps, and provide theoretical assurances on the performance. As an important application, we show how to apply this method to segment microstructure images by considering the images as a point-cloud of image patches. We present results on 2D microscopic images of various materials.