Introduction

This is a Python-based clustering code designed to group similar data points together based on their properties. The algorithm used in this code is Agglomerative Hierarchical clustering. The code takes input data and performs clustering, generating output that can be used for further analysis.

Installation

This code requires the following libraries to be installed:

numpy

matplotlib

scipy

sklearn

You can install these libraries using pip or conda.

Parameters:

data (list): a list of tuples representing the input data points. Each tuple should contain the feature values for one data point. The number of features should match the value of the dimension parameter.

dimension (int): the number of features in the input data points.

method (str): the method used to calculate the distance between clusters when performing hierarchical clustering. Possible values are single, complete, average, weighted, centroid, median, ward.

metric (str): the distance metric used to calculate the distance between individual data points when performing hierarchical clustering. Possible values are euclidean, minkowski, cityblock, cosine, correlation, hamming, jaccard, chebyshev, canberra, braycurtis, mahalanobis.

door (float): a tuning parameter used to determine the number of clusters. The higher the value, the fewer clusters will be generated.

Usage

The code takes input data as a list of tuples containing the coordinates of each data point, and outputs the coordinates of the centroids of each cluster. To use the code, you need to follow these steps:

Store the x and y coordinate information of input data in array x and y.

Update the dimension variable if the input data is multi-dimensional.

Combine the x and y arrays into a list of tuples and store it in the variable data.

Run the code.

Output

The code outputs the coordinates of the centroids of each cluster as a list of lists. Each sublist contains the coordinates of one centroid.