

聚类模型及算法

思考题

- 一．聚类的概念及应用，聚类与分类的区别与联系？
- 二．请分别介绍下面分聚类算法的数学或统计学原理，主要适用的范围，实现的过程或步骤，主要优势和缺陷？
 1. 连通性模型 (Connectivity Models)
 - 1) 近邻传播算法 (Affinity Propagation)
 2. 基于密度的聚类 (Density-Based)
 - 1) DBSCAN 聚类算法 (Density-based spatial clustering of applications with noise)
 - 2) OPTICS 算法 (Ordering Points to Identify the Clustering Structure)
 - 3) 均值漂移聚类算法 (Mean-shift Clustering algorithm)
 3. 基于分布的聚类 (Distribution-based)
 - 1) 高斯混合模型算法 (Gaussian Mixture Model algorithm)
 4. 基于中心的聚类 (Centroid-based)
 - 1) K-均值聚类算法 (K-means clustering algorithm)
 - 2) 小束 K 均值聚类 (Mini-Batch K-means)
 5. 基于层次的聚类 (Hierarchical-based)
 - 1) 分割层次聚类算法 (Divisive Hierarchical Clustering Algorithm)
 - 2) BIRCH-算法 (Balance Iterative Reducing and Clustering

Using Hierarchies algorithm)

3) 凝聚层次聚类算法(Agglomerative hierarchy clustering)

6. 谱聚类(Spectral Clustering)

三. 聚类效果评估方法有哪些? 优势和缺陷是什么?

四. 聚类算法选择原则与方法有哪些?