

0.1 OPTICS 聚类算法

OPTICS (Ordering Points To Identify the Clustering Structure) 是一种基于密度的聚类算法，用于识别任意形状和密度变化的簇。OPTICS 算法通过计算每个数据点的可达距离 (Reachability Distance) 和核心距离 (Core Distance)，并按照可达距离对数据点排序，从而构建簇的层次结构。首先，算法计算每个数据点的核心距离，即它到其邻域内满足最小点数要求的最远距离。然后，计算数据点之间的可达距离，即到达某个点所需的最小核心距离。根据可达距离对数据点进行排序，生成簇的层次结构。

核心距离的计算公式为：

$$\text{core_dist}(p) = \min_{p' \in \text{Neighbors}(p)} \text{dist}(p, p')$$

可达距离的计算公式为：

$$\text{reachability_dist}(o, p) = \max(\text{core_dist}(p), \text{dist}(o, p))$$

其中， $\text{dist}(p, p')$ 是数据点 p 和 p' 之间的距离， $\text{Neighbors}(p)$ 是点 p 的邻域。

OPTICS 算法的优点在于它能够处理任意形状和密度的簇，并且无需预先指定簇的数量。缺点是算法在处理大规模数据集时计算复杂度较高，特别是在高维空间中。OPTICS 算法的时间复杂度为 $O(n \log n)$ ，其中 n 为样本数量。虽然与 DBSCAN 类似，但 OPTICS 能更好地处理具有密度变化的复杂数据集。