

拓扑邻域基在密度聚类算法中的应用

张晓媛^{1,2}, 田毅¹, 任子涵¹, 段天宇¹, 杨斯媛¹, 张月轩¹

(1.河北金融学院统计与数据科学学院, 河北保定 071051; 2.保定市金融大数据智算与应用重点实验室, 河北保定 071051)

摘要:将拓扑学的闭包、聚点、邻域基应用到基于密度的聚类问题中,建立了密度聚类算法的矩阵计算方法,并举例说明了如何通过矩阵乘法使用密度聚类算法对数据集进行聚类。

关键词:密度聚类算法;闭包;聚点;邻域基;矩阵乘法

中图分类号:TP391 **文献标志码:**A

引用格式:张晓媛,田毅,任子涵,等.拓扑邻域基在密度聚类算法中的应用[J].山东大学学报(理学版),2026,61(5):55-64.

Application of topology neighborhood bases in density clustering algorithm

ZHANG Xiaoyuan^{1,2}, TIAN Yi¹, REN Zihan¹, DUAN Tianyu¹, YANG Siyuan¹, ZHANG Yuexuan¹

(1. School of Statistics and Data Science, Hebei Finance University, Baoding 071051, Hebei, China; 2. Baoding Key Laboratory of Financial Big Data Intelligent Computing and Application, Hebei Finance University, Baoding 071051, Hebei, China)

Abstract: Closure, cluster point and neighborhood base in topology are applied to density-based clustering problems. Matrix computation method for density clustering algorithm is proposed, and an example is given to illustrate how to use matrix multiplication to cluster a data set with density clustering algorithm.

Key words: density clustering algorithm; closure; cluster point; neighborhood base; matrix multiplication

0 引言

随着计算机科学技术的发展,拓扑学的应用已经渗透到计算机科学、金融学、生物学、医学、气象学等领域^[1-3]。在数据分析方面,特别是在处理大量复杂的数据时,拓扑学有助于进行数据分析和建模。在人工智能方面,拓扑学可以用于深度学习网络结构的设计和优化、模式识别和图像分类等问题。因此拓扑学在应用领域研究受到越来越多学者的关注。拓扑空间中的邻域、聚点、内部、闭包、边界、连通等概念在许多领域发挥着重要的作用。

在数据挖掘中,聚类是最重要的机器学习算法之一。聚类算法能够发现目标数据集中固有的簇结构,在不知道数据内部关系的情况下,将相似的数据对象聚在一起形成簇,即寻找数据集中彼此非常接近的数据集,因此聚点和邻域在聚类中发挥着举足轻重的作用。文献[4-5]研究了 K -最近邻(K nearest neighbors, KNN)算法。文献[6-8]基于一个点与聚点的距离,研究了用于分类和聚类的 K -均值算法。此后,聚类问题引起了众多学者的极大兴趣^[9-12]。Ester等^[13]研究了基于密度的聚类算法,提出了基于密度的带有噪声的空间聚类应用(density-based spatial clustering of applications with noise, DBSCAN)算法,它不仅能识别球形簇,而且能够正确识别数据集中具有任意大小、形状的簇和噪声,因此得到广泛的研究与应用。本文介绍DBSCAN算法的原理及参数、型拓扑空间和基矩阵的相关概念,将DBSCAN算法的聚类簇与型拓扑空间的传递 p -闭包相联系,利用布尔矩阵的乘法表示拓扑空间中集合的闭包运算。

1 预备知识

1.1 DBSCAN 算法

DBSCAN 算法是一种密度聚类算法,它基于一组邻域参数 (ε, k) 刻画样本分布的紧密程度。假设给定数据集 D , D 中元素 x_{j_1} 的 ε -邻域指包含 D 中与 x_{j_1} 的距离不大于 ε 的数据。若 x_{j_1} 的 ε -邻域至少包含 k 个数据,则 x_{j_1} 是一个核心点;若 x_{j_2} 位于 x_{j_1} 的 ε -邻域中,且 x_{j_1} 是核心点,则称 x_{j_2} 是由 x_{j_1} 密度直达的点;对 x_u 与 x_v ,若存在序列 $\{p_1, p_2, \dots, p_w\} \subseteq D$, 其中 $p_1 = x_u$, $p_w = x_v$, 且 p_{j+1} 由 p_j 密度直达 ($j=1, 2, \dots, w-1$), 则称 x_v 是由 x_u 密度可达的点。若 x 是核心点,则由 x 密度可达的所有数据组成的集合构成聚类簇。不是核心点,但属于某一聚类簇的点称为边界点。既不是核心点,也不是边界点的称为噪声点。图 1 给出了上述概念的直观显示,其中虚线及其内部显示出点的 ε -邻域, x_1 是核心点, x_2 由 x_1 密度直达, x_3 由 x_1 密度可达, x_4 是边界点,不在任一虚线内的点为噪声点^[14]。

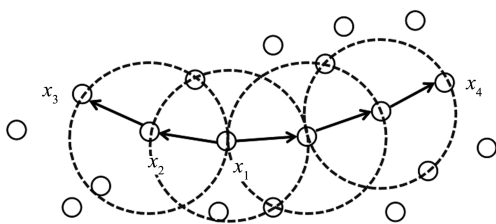


图 1 DBSCAN 定义的基本概念($k=3$)

Fig.1 The basic concepts defined in DBSCAN ($k=3$)

DBSCAN 算法的原理是根据给定的 (ε, k) 找出所有核心点,再以任意核心点为出发点,找出由该核心点密度可达的数据生成聚类簇,直到所有核心点被访问过为止。若 ε 设置过小,可能会导致大部分数据点被认为是噪声点;若 ε 设置过大,可能会将本来属于不同簇的数据合并到一个簇中。 k 的选取通常需要根据数据集的特性和问题需求来选择。较大的 k 可以过滤掉噪声点,但可能会导致较小的簇无法被识别;较小的 k 可以更好地检测小簇,但可能会将噪声点归为一个簇中。

1.2 型拓扑空间

集合 Y 上的一个拓扑 T' 是指满足如下条件的子集族:(1) $\emptyset, Y \in T'$; (2) 若 $U, V \in T'$, 则 $U \cup V \in T'$; (3) 若 $\{U_i; i \in I\} \subseteq T'$, 则 $\bigcup_{i \in I} U_i \in T'$ 。集合 Y 与它上面的拓扑 T' 构成一个拓扑空间 (Y, T') , 简记为 Y 。 T' 中的每个元素称为 Y 的一个开集,开集的补集称为闭集。拓扑空间 Y 的子集 A 称为点 x 的一个邻域,若存在 Y 的开集 U 使得 $x \in U \subseteq A$ 。子集 A 的内部 A° 是指包含在 A 中的所有开集的并, A 的闭包 \bar{A} 指包含 A 的所有闭集的交, $\bar{A} - A^\circ$ 是 A 的边界。 $x \in \bar{A}$ 当且仅当每一个包含 x 的开集 U , 都有 $U \cap A \neq \emptyset$ 。如果 x 的任意一个邻域 V , $V \cap A \setminus \{x\} \neq \emptyset$, 即 x 属于 $A \setminus \{x\}$ 的闭包, 则称 x 为 A 的一个聚点。更多关于拓扑学的相关内容可参看文献[15]。

ε -邻域是 DBSCAN 算法中非常重要的参数之一,考虑将拓扑学中邻域的概念变为带有某种“半径”的邻域。文献[16]在拓扑空间 (Y, T') 和偏序集 (P', \leq) 中引入了型拓扑空间,并定义了点的 ε 型开邻域和子集的 ε -聚点、直接闭包和传递 ε -闭包。首先介绍偏序集的概念。设 \leq 是集合 P' 上的二元关系,若 \leq 满足下列条件,则称 \leq 是 P' 上的一个偏序, (P', \leq) 称为偏序集:(1) 自反性,对任意 $a \in P'$, $a \leq a$ 恒成立;(2) 传递性,对任意 $a, b, c \in P'$, 若 $a \leq b$, $b \leq c$, 则 $a \leq c$; (3) 反对称性,对任意 $a, b \in P'$, 若 $a \leq b$ 且 $b \leq a$, 则 $a = b$ 。在不引起混淆时, (P', \leq) 简记为 P' , 比如,实数集按照通常的大小关系构成一个偏序集。

定义 1^[16] 设 (Y, T') 是拓扑空间, (P', \leq) 是偏序集, $x \in Y$, 记 $U(x) = \{U \in T' : x \in U\}$, 即 $U(x)$ 是 x 的所有开邻域构成的集族。若对任意 $x \in Y$, 存在函数 $\sigma_x : U(x) \rightarrow P'$ 满足对任意 $U, V \in U(x)$, 当 $U \subseteq V$ 时, $\sigma_x(U) \leq \sigma_x(V)$, 则称五元组 $(Y, T', P', \leq, \{\sigma_x : x \in Y\})$ 为型拓扑空间。当 $\sigma_x(U) = \varepsilon$ 时, 称 U 是 x 的 ε 型开邻域。 ε - $U(x)$ 表示 U 是 x 的 ε 型开邻域。

对任意 $x \in Y$, 假设 $\varepsilon-U(x)$ 是 x 的 ε 型开邻域, 若对 x 的任一 ε 型开邻域 $\varepsilon-V(x)$, 都有 $U \subseteq V$ 成立, 则称 σ 有最小的 ε 型开邻域。当 Y 为有限集时, 对任意 $\varepsilon \in P$, σ 有最小的 ε 型开邻域。

例 1 设 X 为有限集, T 为 X 上的欧氏拓扑, P 为正实数赋予通常的线性序, 对每个 $x \in X$ 和 $\varepsilon \in P$, 令 $U_\varepsilon(x) = \{y \in X: d(x, y) \leq \varepsilon\}$, 其中 $d(x, y)$ 为 x 到 y 的距离。定义函数 $\sigma_x(U_\varepsilon(x)) = \varepsilon$, 则生成了一个型拓扑空间。下文中所指的型拓扑空间 $(X, T, P, \leq, \{\sigma_x: x \in X\})$ 总是这样生成的, x 的 ε 型开邻域 $\varepsilon-U(x)$ 总是指 $U_\varepsilon(x) = \{y \in X: d(x, y) \leq \varepsilon\}$ 。

定义 2^[16] 设 $(X, T, P, \leq, \{\sigma_x: x \in X\})$ 为型拓扑空间, $A \subseteq X$, $\varepsilon \in P$, 点 $x \in X$ 称为 A 的 ε -聚点, 若对任意 $\varepsilon-U(x)$, 都有 $U \cap A \neq \emptyset$ 。 $A \cup \{x \in X: x \text{ 是 } A \text{ 的 } \varepsilon\text{-聚点}\}$ 称为 A 的直接闭包, 记为 $\varepsilon\text{-CL}_1(A)$ 。

注 1 在一般拓扑空间中, $\overline{(\overline{A})} = \overline{A}$, 但是在型拓扑空间中, $\varepsilon\text{-CL}_1(\varepsilon\text{-CL}_1(A))$ 一般不等于 $\varepsilon\text{-CL}_1(A)$ 。因为若 x 是 $\varepsilon\text{-CL}_1(A)$ 的 ε -聚点, 对任意 $\varepsilon-U(x)$, $U \cap (\varepsilon\text{-CL}_1(A)) \neq \emptyset$, 假设 $y \in U \cap (\varepsilon\text{-CL}_1(A))$, 但是 U 未必是 y 的 ε 型开邻域, 所以不一定有 $U \cap A \neq \emptyset$, 故 x 不一定是 A 的 ε -聚点。记 $\varepsilon\text{-CL}_2(A) = \varepsilon\text{-CL}_1(\varepsilon\text{-CL}_1(A))$, $\varepsilon\text{-CL}_3(A) = \varepsilon\text{-CL}_1(\varepsilon\text{-CL}_2(A))$, \dots , $\varepsilon\text{-CL}_n(A) = \varepsilon\text{-CL}_1(\varepsilon\text{-CL}_{n-1}(A))$ 。

A 的传递 ε -闭包定义为 $\varepsilon\text{-tr}(A) = \cup \{\varepsilon\text{-CL}_n(A): n \in \mathbf{N}\}$ 。当 $A = \{x\}$ 时, $\varepsilon\text{-CL}_1(x)$ 表示 $\varepsilon\text{-CL}_1(\{x\})$, $\varepsilon\text{-tr}(x)$ 表示 $\varepsilon\text{-tr}(\{x\})$ 。

例 2 设 $E = \{(0, 0), (1, 0), (1, 1), (2, 2)\}$ 为平面上的 4 个点赋予欧氏拓扑, P 为正实数赋予通常的线性序, 对任意 $x \in E$ 和 $\varepsilon \in P$, $U_\varepsilon(x)$ 和函数 σ_x 为例 1 中所定义的。若 ε 是大于 1 小于 $\sqrt{2}$ 的任一实数, 则

$$\begin{aligned} U_\varepsilon((0, 0)) &= \{(0, 0), (1, 0)\}, & U_\varepsilon((1, 0)) &= \{(0, 0), (1, 0), (1, 1)\}, \\ U_\varepsilon((1, 1)) &= \{(1, 1), (1, 0)\}, & U_\varepsilon((2, 2)) &= \{(2, 2)\}. \\ \varepsilon\text{-tr}((1, 0)) &= \varepsilon\text{-CL}_1((1, 0)) = \{(0, 0), (1, 0), (1, 1)\}, \\ \varepsilon\text{-tr}((2, 2)) &= \varepsilon\text{-CL}_1((2, 2)) = \{(2, 2)\}. \end{aligned}$$

$\varepsilon\text{-CL}_1((0, 0)) = \{(0, 0), (1, 0)\}$, 对任意 $n \geq 2$, $\varepsilon\text{-CL}_n((0, 0)) = \{(0, 0), (1, 0), (1, 1)\}$, 因此, $\varepsilon\text{-tr}((0, 0)) = \{(0, 0), (1, 0), (1, 1)\}$ 。类似可得 $\varepsilon\text{-tr}((1, 1)) = \{(0, 0), (1, 0), (1, 1)\}$ 。

定义 3^[16] 设 $(X, T, P, \leq, \{\sigma_x: x \in X\})$ 为型拓扑空间, 对任意 $x, y \in X$, 若存在 x 的 ε 型开邻域 $\varepsilon-U(x)$ 满足 $y \notin (\varepsilon-U(x))$, 则存在 y 的 ε 型开邻域 $\varepsilon-V(y)$ 使得 $x \notin (\varepsilon-V(y))$, 称 $(X, T, P, \leq, \{\sigma_x: x \in X\})$ 为 ε -对称型的。例 2 中的型拓扑空间是对称型的。

定理 1^[16] 设 $(X, T, P, \leq, \{\sigma_x: x \in X\})$ 为 ε -对称型的型拓扑空间, 假设 σ 有最小的 ε 型开邻域, 则对任意 $x \in X$, 任意 $y \in (\varepsilon\text{-tr}(x))$, 都有 $\varepsilon\text{-tr}(x) = \varepsilon\text{-tr}(y)$ 。

1.3 基矩阵

本节主要介绍拓扑空间中基和局部基的概念, 以及如何通过布尔矩阵的乘法表示拓扑空间中一个集合的闭包运算。

定义 4^[11] 设 (Y, T') 是拓扑空间, $B \subseteq T'$ 满足对 Y 的每个开集 U 和 U 中元素 x , 至少存在 B 中的一个元素 C , 使得 $x \in C \subseteq U$, 则称 B 是拓扑空间 (Y, T') 的一个基。 $B(x) = \{C \in B: x \in C\}$ 称为点 x 的局部基。

定义 5^[17] 若矩阵中每个元素为 0 或 1, 则称矩阵为布尔矩阵。假设 $A = (a_{de})_{z \times m}$, $B = (b_{ef})_{m \times l}$ 是两个布尔矩阵, 乘积矩阵 $C = AB = (c_{df})_{z \times l}$ 定义为

$$c_{df} = \bigvee_{e=1}^m (a_{de} \wedge b_{ef}), \tag{1}$$

式中, $d = 1, 2, \dots, z$, $f = 1, 2, \dots, l$, \bigvee 和 \wedge 分别为取大和取小运算。

例如, 布尔矩阵 $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$, 则乘积矩阵 $C = AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ 。

定义 6^[17] 设 $Z = \{x_1, x_2, \dots, x_{n_1}\}$ 是有限个元素构成的集合, 其中元素的序已经给定, Z' 是 Z 的任一子集, 称 $X_{Z'} = (a_1, a_2, \dots, a_{n_1})$ 为子集 Z' 的特征函数向量, 其中 $a_r = \begin{cases} 1, & x_r \in Z' \\ 0, & x_r \notin Z' \end{cases}$, $r = 1, 2, \dots, n_1$ 。

例如,若 $G = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, 则 G 的子集 $G' = \{x_1, x_2, x_5\}$ 的特征函数向量 $X_{G'} = (1, 1, 0, 0, 1, 0, 0)$ 。

定义 7^[17] 设 $W = \{x_1, x_2, \dots, x_k\}$ 是序已经给定的有限个元素构成的集合, T 是 W 上的拓扑, 若布尔矩

阵 $M_{k \times k} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{pmatrix}$ 满足每一行 α_ζ 是点 x_ζ 局部基的特征函数向量, 即 $X_{B(x_\zeta)} = \alpha_\zeta$, 则称布尔矩阵 $M_{k \times k}$ 是拓扑空间

(W, T) 的基矩阵。

注 2 设 $R = \{x_1, x_2, \dots, x_r\}$ 是序已经给定的有限个元素构成的集合, 则有以下结论:

(1) 若 T 是 R 上的拓扑, 则基矩阵的第 h 行是 T 中含有 x_h 的最小开集的特征函数向量。

(2) R 上的拓扑与基矩阵之间是一一对应的。

定理 2^[17] 设 X 是序已经给定的有限个元素构成的集合, T 是 X 上的拓扑, M 是拓扑空间 (X, T) 的基矩阵, 则对 X 的任一子集 A , \bar{A} 的特征函数向量等于基矩阵与 A 的特征函数向量的乘积, 即 $X_{\bar{A}} = MX_A$ 。

2 DBSCAN 算法的矩阵计算

将型拓扑空间中的直接闭包和传递闭包应用于 DBSCAN 算法, 借助于拓扑空间中集合闭包的矩阵运算, 将矩阵乘法应用到 DBSCAN 算法中对数据集进行聚类。

定理 3 设 $(X, T, P, \leq, \{\sigma_x : x \in X\})$ 为型拓扑空间, 在 DBSCAN 算法中给定一组 (ε, k) , 若 x 是核心点, y 是由 x 密度可达的点, 则 $y \in \varepsilon\text{-tr}(x)$; 反之, 若 $y \in \varepsilon\text{-tr}(x)$, 且存在核心点 q_0, q_1, \dots, q_ν , 其中 $q_0 = x$, $q_\nu = y$ 满足对任意 $s \in \{0, 1, \dots, \nu-1\}$, $q_{s+1} \in (\varepsilon\text{-CL}_1(q_s))$, 则 y 是由 x 密度可达的点。因此由 x 生成的聚类簇为 $\varepsilon\text{-tr}(x)$ 。

证明 首先证明 b 是由 a 直达的点当且仅当 $b \in (\varepsilon\text{-CL}_1(a))$ 。事实上, b 是由 a 直达的点当且仅当 b 位于 a 的 ε -邻域中, 即 $d(a, b) \leq \varepsilon$ 当且仅当 $d(b, a) \leq \varepsilon$, 即 a 也位于 b 的 ε -邻域中, 故 $b \in (\varepsilon\text{-CL}_1(a))$ 。

假设 y 是由 x 密度可达的点, 则存在 p_1, p_2, \dots, p_w , 其中 $p_1 = x$, $p_w = y$, 且 p_{j+1} 由 p_j 密度直达。由以上证明可得, 对任意 $j = 1, 2, \dots, w-1$, $p_{j+1} \in (\varepsilon\text{-CL}_1(p_j))$, 从而 $y \in \varepsilon\text{-tr}(x)$ 。反之, 若 $y \in \varepsilon\text{-tr}(x)$, 且存在核心点 q_0, q_1, \dots, q_ν , 其中 $q_0 = x$, $q_\nu = y$, 满足对任意 $s \in \{0, 1, \dots, \nu-1\}$, $q_{s+1} \in (\varepsilon\text{-CL}_1(q_s))$, 则 q_{s+1} 由 q_s 密度直达, 从而 y 由 x 密度可达。根据 DBSCAN 算法原理, 寻找一个核心点生成的聚类簇, 就是要找到与该核心点密度可达的所有点, 所以由 x 生成的聚类簇为 $\cup \{\varepsilon\text{-CL}_n(x) : n \in \mathbf{N}\} = \varepsilon\text{-tr}(x)$ 。

类似于拓扑空间中基矩阵的定义, 我们给出型拓扑空间中 p -基矩阵的定义。

定义 8 设五元组 $(X, T, P, \leq, \{\sigma_x : x \in X\})$ 为 m' 个元素构成的型拓扑空间, 在 DBSCAN 算法中给定一组 (ε, k) , 记核心点的集合为 Ω , 称向量 $X_{\varepsilon\text{-}U(x_o)} = (b_1, b_2, \dots, b_{m'})$ 为点 x_o 的 ε 型开邻域的特征函数向量, 其中 $b_\sigma (\sigma = 1, 2, \dots, m')$ 定义为

(1) 如果 x_o 是核心点, 定义 $b_\sigma = \begin{cases} 1, & x_o \in \varepsilon\text{-}U(x_o), \\ 0, & x_o \notin \varepsilon\text{-}U(x_o), \end{cases}$

(2) 如果 x_o 不是核心点, 定义 $b_\sigma = \begin{cases} 1, & x_o \in \Omega \text{ 且 } x_o \in \varepsilon\text{-}U(x_o) \text{ 或 } \sigma = 0, \\ 0, & \text{否则。} \end{cases}$

定义 9 设五元组 $(X, T, P, \leq, \{\sigma_x : x \in X\})$ 为 n 个元素构成的型拓扑空间, 若布尔矩阵 $M'_{n \times n} =$

$\begin{pmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_n \end{pmatrix}$ 满足每一行 α'_i 是点 x_i 的 ε 型开邻域的特征函数向量, 即 $X_{\varepsilon\text{-}U(x_i)} = \alpha'_i$, 则称布尔矩阵 $M'_{n \times n}$ 是 X 的 ε -基

矩阵。

定理 4 设五元组 $(X, T, P, \leq, \{\sigma_x : x \in X\})$ 为 n 个元素构成的型拓扑空间, M' 是 X 的 ε -基矩阵。在 DBSCAN 算法中给定一组 (ε, k) , 则对 X 的任一核心点 x ,

$$X_{\varepsilon\text{-CL}_{g+1}(x)} = M' X_{\varepsilon\text{-CL}_g(x)}, \quad (2)$$

式中, $g=0, 1, \dots, X_{\varepsilon\text{-CL}_0(x)} = X_{|x|}$ 。

证明 假设 $M' X_{\varepsilon\text{-CL}_g(x)} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$, $c_\zeta = 1$, 分以下 2 种情况来证明。

(1) 若 x_ζ 是核心点,

$$\begin{aligned} c_\zeta = 1 &\Leftrightarrow \text{存在 } k' \text{ 使得 } X_{\varepsilon\text{-}U(x_{k'})}(x_{k'}) = 1 \text{ 且 } X_{\varepsilon\text{-CL}_g(x)}(x_{k'}) = 1 \\ &\Leftrightarrow \text{存在 } k' \text{ 使得 } x_{k'} \in \varepsilon\text{-}U(x_\zeta) \text{ 且 } x_{k'} \in \varepsilon\text{-CL}_g(x) \\ &\Leftrightarrow \varepsilon\text{-}U(x_\zeta) \cap \varepsilon\text{-CL}_g(x) \neq \emptyset \\ &\Leftrightarrow x_\zeta \in \varepsilon\text{-CL}_1(\varepsilon\text{-CL}_g(x)) = \varepsilon\text{-CL}_{g+1}(x) \\ &\Leftrightarrow X_{\varepsilon\text{-CL}_{g+1}(x)}(x_\zeta) = 1。 \end{aligned}$$

因此, $X_{\varepsilon\text{-CL}_{g+1}(x)} = M' X_{\varepsilon\text{-CL}_g(x)}$ 。

(2) 若 x_ζ 不是核心点,

$$\begin{aligned} c_\zeta = 1 &\Leftrightarrow \text{存在 } s' (s' \neq \zeta) \text{ 使得 } x_{s'} \text{ 是核心点, } x_\zeta \in \varepsilon\text{-}U(x_{s'}) \text{ 且 } X_{\varepsilon\text{-CL}_g(x)}(x_{s'}) = 1 \text{ 或 } x_\zeta \in \varepsilon\text{-CL}_g(x) \\ &\Leftrightarrow \text{存在 } s' (s' \neq \zeta) \text{ 使得 } x_{s'} \text{ 是核心点, } x_{s'} \in \varepsilon\text{-}U(x_\zeta) \text{ 且 } X_{\varepsilon\text{-CL}_g(x)}(x_{s'}) = 1 \text{ 或 } x_\zeta \in \varepsilon\text{-CL}_g(x) \\ &\Leftrightarrow \text{存在 } s' (s' \neq \zeta) \text{ 使得 } x_{s'} \text{ 是核心点, } x_{s'} \in \varepsilon\text{-}U(x_\zeta) \text{ 且 } x_{s'} \in \varepsilon\text{-CL}_g(x) \text{ 或 } x_\zeta \in \varepsilon\text{-CL}_g(x) \\ &\Leftrightarrow \varepsilon\text{-}U(x_\zeta) \cap \varepsilon\text{-CL}_g(x) \neq \emptyset \\ &\Leftrightarrow x_\zeta \in \varepsilon\text{-CL}_1(\varepsilon\text{-CL}_g(x)) = \varepsilon\text{-CL}_{g+1}(x) \\ &\Leftrightarrow X_{\varepsilon\text{-CL}_{g+1}(x)}(x_\zeta) = 1。 \end{aligned}$$

因此, $X_{\varepsilon\text{-CL}_{g+1}(x)} = M' X_{\varepsilon\text{-CL}_g(x)}$ 。

根据定理 3, 4, 按以下步骤实现 DBSCAN 算法的矩阵计算。

步骤 1 输入数据集 $X = \{x_1, x_2, \dots, x_n\}$, 固定 X 中元素的顺序, 输入 (ε, k) 。

步骤 2 根据 (ε, k) 确定 Ω 。

步骤 3 在给定的 X 的顺序下, 计算 X 的 ε -基矩阵。

步骤 4 对每个核心对象 x , 根据式(2)计算 $X_{\varepsilon\text{-CL}_1(x)}, X_{\varepsilon\text{-CL}_2(x)}, \dots, X_{\varepsilon\text{-CL}_g(x)}, \dots$, 若 $X_{\varepsilon\text{-CL}_g(x)} = X_{\varepsilon\text{-CL}_{g+1}(x)}$, 则终止计算, 记 $C(x) = \varepsilon\text{-tr}(x) = \varepsilon\text{-CL}_g(x)$ 。

步骤 5 输出簇划分 $C = \{C(x) : x \in \Omega\}$ 。

例 3 设数据集 Q 为平面直角坐标系中的点 $(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (2, 2), (2, 3), (3, 3), (4, 4)$ 。为方便表示, 记 $a_1 = (0, 0), a_2 = (1, 0), a_3 = (2, 0), a_4 = (0, 1), a_5 = (1, 1), a_6 = (2, 2), a_7 = (2, 3), a_8 = (3, 3), a_9 = (4, 4)$, 在 X 上赋予欧氏拓扑。定义 $U_\varepsilon(x) = \{y \in Q : d(x, y) \leq \varepsilon\}$, 其中 $d(x, y)$ 为 x 到 y 的距离, $\sigma_x(U_\varepsilon(x)) = \varepsilon$, 则该型拓扑空间有最小的 ε 型开邻域且为 ε -对称型的。由于数据集较少, 对参数分别取不同值。

(1) 选取 $\varepsilon=1, k=3$ 的聚类。

步骤 1 固定 Q 中元素的顺序 $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$, 取 $\varepsilon=1, k=3$ 。

步骤 2 根据 (ε, k) , 确定 $\Omega_{1,3} = \{a_1, a_2, a_4, a_5, a_7\}$ 。

步骤 3 计算 ε -基矩阵。以 a_1 为核心点, 根据定义 8(1) 计算点 a_1 的 ε 型开邻域的特征函数向量为 $\alpha_1 = (1, 1, 0, 1, 0, 0, 0, 0, 0)$, 用同样的方法分别计算核心点 a_2, a_4, a_5, a_7 的 ε 型开邻域的特征函数向量;

a_3 不是核心点,根据定义 8(2) 计算点 a_3 的 ε 型开邻域的特征函数向量为 $\alpha_3 = (0, 1, 1, 0, 0, 0, 0, 0, 0)$, 用同样的方法分别计算非核心点 a_6, a_8, a_9 的 ε 型开邻域的特征函数向量, 得到 X 的 ε -基矩阵为

$$M_{1,3} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

步骤 4 根据定理 4 计算核心对象的传递 ε -闭包为

$$\begin{aligned} X_{\varepsilon\text{-CL}_1(a_1)}^1 &= M_{1,3} X_{\{a_1\}} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \\ X_{\varepsilon\text{-CL}_2(a_1)}^1 &= M_{1,3} X_{\varepsilon\text{-CL}_1(a_1)}^1 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \\ X_{\varepsilon\text{-CL}_3(a_1)}^1 &= M_{1,3} X_{\varepsilon\text{-CL}_2(a_1)}^1 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

因此,核心对象 a_1 生成的聚类簇为 $C_1(a_1) = \varepsilon\text{-tr}(a_1) = \varepsilon\text{-CL}_2(a_1) = \{a_1, a_2, a_3, a_4, a_5\}$ 。

由于该型拓扑空间有最小的 ε 型开邻域且为 ε -对称型的,因此 $\varepsilon\text{-tr}(a_1) = \varepsilon\text{-tr}(a_2) = \varepsilon\text{-tr}(a_4) = \varepsilon\text{-tr}(a_5)$ 。

$$\begin{aligned}
\mathbf{X}_{\varepsilon\text{-CL}_1(a_7)}^1 &= \mathbf{M}_{1,3} \mathbf{X}_{|a_7|} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \\
\mathbf{X}_{\varepsilon\text{-CL}_2(a_7)}^1 &= \mathbf{M}_{1,3} \mathbf{X}_{\varepsilon\text{-CL}_1(a_7)}^1 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}.
\end{aligned}$$

因此,核心对象 a_7 生成的聚类簇为 $C_1(a_7) = \varepsilon\text{-tr}(a_7) = \varepsilon\text{-CL}_1(a_7) = \{a_6, a_7, a_8\}$ 。

步骤 5 输出聚类簇的结果。最终得到 2 个聚类簇 $C_1(a_1) = \{a_1, a_2, a_3, a_4, a_5\}$ 、 $C_1(a_7) = \{a_5, a_6, a_7\}$ 。边界点为 $\{a_3, a_6, a_8\}$, 噪声点为 $\{a_9\}$ 。

(2) 选取 $\varepsilon=1, k=4$ 的聚类。

步骤 1 固定 Q 中元素的顺序 $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$, 取 $\varepsilon=1, k=4$ 。

步骤 2 根据 (ε, k) , 确定 $\Omega_{1,4} = \{x_2\}$ 。

$$\text{步骤 3 计算 } \varepsilon\text{-基矩阵 } \mathbf{M}_{1,4} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

步骤 4 计算核心对象的传递 ε -闭包为

$$\mathbf{X}_{\varepsilon\text{-CL}_1(a_2)}^2 = \mathbf{M}_{1,4} \mathbf{X}_{|a_2|} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$X_{\varepsilon\text{-CL}_2(a_2)}^2 = M_{1,4} X_{\varepsilon\text{-CL}_1(a_2)}^2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}。$$

因此,核心对象 a_2 生成的聚类簇为 $C_2(a_2) = \varepsilon\text{-tr}(a_2) = \varepsilon\text{-CL}_1(a_2) = \{a_1, a_2, a_3, a_5\}$ 。

步骤 5 输出聚类簇的结果。最终得到 1 个聚类 $C_2(a_2) = \{a_1, a_2, a_3, a_5\}$ 。边界点为 $\{a_1, a_3, a_5\}$, 噪声点为 $\{a_4, a_6, a_7, a_8, a_9\}$ 。

(3) 选取 $\varepsilon = \sqrt{2}$, $k=4$ 的聚类。

步骤 1 固定 Q 中元素的顺序 $\{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$, 取 $\varepsilon = \sqrt{2}$, $k=4$ 。

步骤 2 根据 (ε, k) , 确定 $\Omega_{\sqrt{2}, 4} = \{a_1, a_2, a_4, a_5, a_6, a_8\}$ 。

步骤 3 计算 ε -基矩阵 $M_{\sqrt{2}, 4} =$

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}。$$

步骤 4 计算核心对象的传递 ε -闭包为

$$X_{\varepsilon\text{-CL}_1(a_1)}^3 = M_{\sqrt{2}, 4} X_{\{a_1\}}^3 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$X_{\varepsilon\text{-CL}_2(a_1)}^3 = M_{\sqrt{2}, 4} X_{\varepsilon\text{-CL}_1(a_1)}^3 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{aligned}
 X_{\varepsilon\text{-CL}_3(a_1)}^3 &= M_{\sqrt{2},4} X_{\varepsilon\text{-CL}_2(a_1)}^3 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \\
 X_{\varepsilon\text{-CL}_4(a_1)}^4 &= M_{\sqrt{2},4} X_{\varepsilon\text{-CL}_3(a_1)}^4 = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}.
 \end{aligned}$$

因此,核心对象 a_1 生成的聚类簇为 $C_3(a_1) = \varepsilon\text{-tr}(a_1) = \varepsilon\text{-CL}_4(a_1) = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$ 。

步骤 5 输出聚类簇的结果。最终得到 1 个聚类簇 $C_3(a_1) = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$, 边界点为 $\{a_3, a_7, a_9\}$, 无噪声点。

3 结语

本文将型拓扑空间中的聚点和传递闭包的概念应用到 DBSCAN 算法中,通过布尔矩阵乘法表示拓扑空间中集合闭包运算的方法,给出了 DBSCAN 算法的矩阵计算方法,为 DBSCAN 算法提供了一种新的计算手段。

参考文献:

[1] 曹雨康. 基于拓扑学和机器学习的药物属性预测研究[D]. 武汉:武汉纺织大学,2022.
CAO Yukang. Research on drug properties prediction based on topology and machine learning[D]. Wuhan: Wuhan Textile University, 2022.

[2] 傅秋婷. 基于拓扑数据分析技术的股票价格时间序列聚类研究[D]. 成都:西南财经大学,2022.
FU Qiuting. Research on clustering of stock price time series based on topological data analysis[D]. Chengdu: Southwestern University of Finance and Economics, 2022.

[3] 张家琛. 基于拓扑数据分析方法的人体步态研究[D]. 天津:天津理工大学,2022.
ZHANG Jiachen. Research on human gait based on topological data analysis[D]. Tianjin: Tianjin University of Technology, 2022.

[4] FIX E, HODGES J L. Discriminatory analysis, nonparametric discrimination: consistency properties[R]. Texas: USAF School of Aviation Medicine, 1951.

[5] COVER T M, HART P E. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1):21-27.

[6] MACQUEEN J. Some methods for classification and analysis of multivariate observations[J]. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967, 5(1):281-297.

- [7] FORGY E W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications[J]. *Biometrics*, 1965, 21(3):768-769.
- [8] LLOYD S. Least square quantization in PCM[J]. *IEEE Transactions on Information Theory*, 1982, 28(2):129-137.
- [9] JOHNSON S. Hierarchical clustering schemes[J]. *Psychometrika*, 1967, 32(3):241-254.
- [10] HARDING E F. The number of partitions of a set of n points in k dimensions induced by hyperplanes[J]. *Proceedings of the Edinburgh Mathematical Society*, 1967, 15(4):285-289.
- [11] JARDINE N, SIBSON R. The construction of hierarchic and non-hierarchic classifications[J]. *The Computer Journal*, 1968, 11(2):117-184.
- [12] HARTIGAN J A. Clustering algorithms[M]. Toronto: Wiley, 1975.
- [13] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C] // *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. Oregon: AAAI, 1996.
- [14] 周志华. 机器学习[M]. 北京:清华大学出版社,2016.
ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016.
- [15] ENGELKING R. General topology[M]. Berlin: Heldermann Verlag, 1989.
- [16] HU Wanjun. Typed topology and its application to data sets[J]. *Topology and Its Applications*, 2024, 342:108760.
- [17] MA Liwen. Important matrix computations in finite topological spaces[J]. *Applied Mathematics and Computation*, 2021, 395:125808.

(编辑:陈丽萍)

(上接第 54 页)

- [18] 李翔宇,张雪芹. ORBTSDF-SCNet:一种动态场景在线三维重建方法[J]. *华东理工大学学报(自然科学版)*, 2023, 49(2):284-294.
LI Xiangyu, ZHANG Xueqin. ORBTSDF-SCNet: an online 3D reconstruction method for dynamic scene[J]. *Journal of East China University of Science and Technology*, 2023, 49(2):284-294.
- [19] 史西兵,赵政文. 移动物联网负载终端坐标动态转换算法[J]. *计算机仿真*, 2021, 38(1):247-250.
SHI Xibing, ZHAO Zhengwen. Dynamic coordinate transformation algorithm of mobile internet of things load terminal[J]. *Computer Simulation*, 2021, 38(1):247-250.
- [20] 田野,陈宏巍,王法胜,等. 室内移动机器人的 SLAM 算法综述[J]. *计算机科学*, 2021, 48(9):223-234.
TIAN Ye, CHEN Hongwei, WANG Fasheng, et al. Overview of SLAM algorithm for mobile robots[J]. *Computer Science*, 2021, 48(9):223-234.
- [21] 程保喜. GNSS 与惯性导航组合系统在复杂环境下的定位研究[J]. *中北大学学报(自然科学版)*, 2021, 42(1):89-96.
CHENG Baoxi. Research on positioning of GNSS and inertial navigation integrated system in complex environment [J]. *Journal of North University of China (Natural Science Edition)*, 2021, 42(1):89-96.
- [22] 李帅鑫,李广云,周阳林,等. 改进的单目视觉实时定位与测图方法[J]. *仪器仪表学报*, 2017, 38(11):2849-2857.
LI Shuaxin, LI Guangyun, ZHOU Yanglin, et al. Improved monocular simultaneous localization and mapping solution[J]. *Chinese Journal of Scientific Instrument*, 2017, 38(11):2849-2857.
- [23] 祝会忠,范佳宝,李军. 低成本 GNSS 终端的抗差 RTK 定位性能研究[J]. *测绘科学*, 2023, 48(12):1-9.
ZHU Huizhong, FAN Jiabao, LI Jun. Research on the positioning performance of robust RTK low-cost GNSS terminals [J]. *Science of Surveying and Mapping*, 2023, 48(12):1-9.

(编辑:陈丽萍)