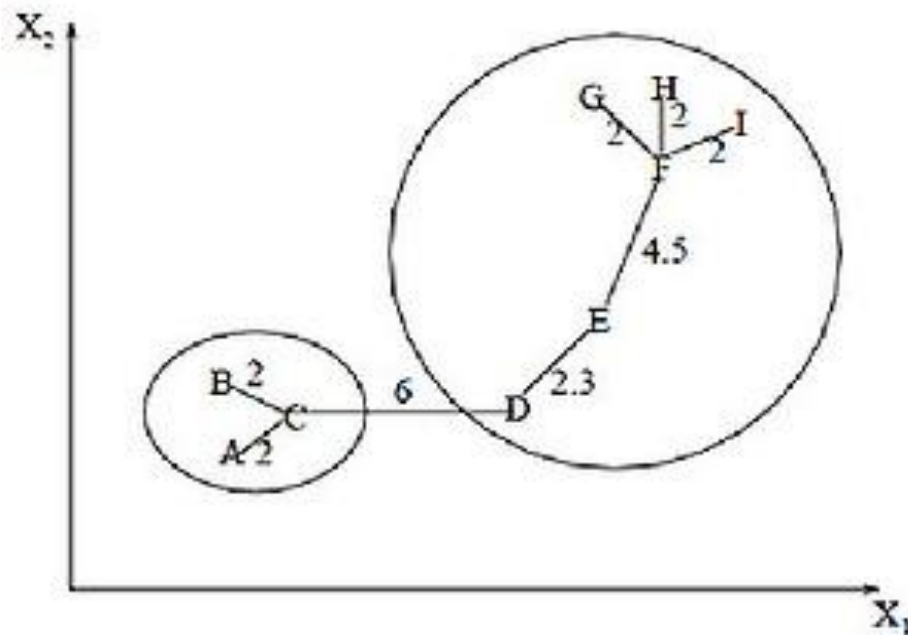


Послойная кластеризация

Выполнила: Шамсутдинова Лилия

Алгоритмы, основанные на теории графов

- Выборка объектов представляется в виде графа $G=(V,E)$



Алгоритм послойной кластеризации

- Основан на выделении связных компонент графа на некотором уровне расстояний между объектами (вершинами)

- С-порог расстояния

$$0 \leq \rho(x, x') \leq 1 \quad 0 \leq c \leq 1$$

- Формирует последовательность подграфов графа G , которые отражают иерархические связи между кластерами

$$G^0 \subseteq G^1 \subseteq \dots \subseteq G^m$$

Алгоритм послойной кластеризации

- $G^t = (V, E^t)$ — граф на уровне c^t

$$E^t = \{e_{ij} \in E : \rho_{ij} \leq c_t\}$$

- c^t — t -ый порог расстояния
- m — количество уровней иерархии
- $G^0 = (V, o)$, o — пустое множество ребер графа, получаемое при $c^0 = 0$
- $G^m = G$, то есть граф объектов без ограничений на расстояние (длину ребер графа), поскольку $c^m = 1$

Алгоритм послойной кластеризации

- Посредством изменения порогов расстояния $\{c^0, \dots, c^m\}$, где $0 = c^0 < c^1 < \dots < c^m = 1$, возможно контролировать глубину иерархии получаемых кластеров
- Вычислительная сложность: $O(\max(n, m))$, где $m < n(n-1)/2$
- Входные данные: последовательность порогов расстояния
- Результат: древовидная структура кластеров с разными уровнями иерархии

Практика

- Даны объекты:

$$\begin{pmatrix} 1 \\ 5 \\ 4 \\ 7 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 4 \\ 4 \\ 5 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 3 \\ 9 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 3 \\ 1 \\ 3 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 6 \\ 7 \\ 8 \\ 9 \\ 0 \end{pmatrix}$$

- В качестве расстояния использовать расстояние Чебышева: $l_{\infty}(\vec{x}, \vec{y}) = \max_{i=1, \dots, n} |x_i - y_i|$
- $C1 = 4$, $C2 = 5$
- Представить решение в виде графа