**Лабораторная работа 1**

**Кластеризация**

**Пример.** Рассмотрим на примере один из алгоритмов кластеризации ­– агломеративный иерархический алгоритм.

Пусть есть шесть объектов. Все попарные расстояния между ними уже посчитаны. Необходимо их разбить на кластеры.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | **4** | **5** | 6 |
| 1 | 0 | 2.83 | 3.16 | 10.2 | 12.17 | 13.6 |
| 2 | 2.83 | 0 | 3.16 | 8.94 | 10.77 | 12.53 |
| 3 | 3.16 | 3.16 | 0 | 7.07 | 9.06 | 10.44 |
| **4** | 10.2 | 8.94 | 7.07 | 0 | 2 | 3.61 |
| **5** | 12.17 | 10.77 | 9.06 | 2 | 0 | 2.24 |
| 6 | 13.6 | 12.53 | 10.44 | 3.61 | 2.24 | 0 |

Считаем вначале, что каждый объект ­– это отдельный кластер. Выбираем два ближайших кластера и сливаем их в один. Это оказались кластеры 4 и 5.

Находим расстояние от всех остальных кластеров до нового кластера 4,5 по принципу «ближайшего соседа». Например, расстояние от 1 до 4,5 равно 10.2, так как ближайшие точки этих кластеров – это 1 и 4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | **4,5** | **6** |
| 1 | 0 | 2.83 | 3.16 | 10.2 | 13.6 |
| 2 | 2.83 | 0 | 3.16 | 8.94 | 12.53 |
| 3 | 3.16 | 3.16 | 0 | 7.07 | 10.44 |
| **4,5** | 10.2 | 8.94 | 7.07 | 0 | 2.24 |
| **6** | 13.6 | 12.53 | 10.44 | 2.24 | 0 |

Снова берём два ближайших кластера и объединяем их. Это кластеры 4,5 и 6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **1** | **2** | 3 | 4,5,6 |
| **1** | 0 | 2.83 | 3.16 | 10.2 |
| **2** | 2.83 | 0 | 3.16 | 8.94 |
| 3 | 3.16 | 3.16 | 0 | 7.07 |
| 4,5,6 | 10.2 | 8.94 | 7.07 | 0 |

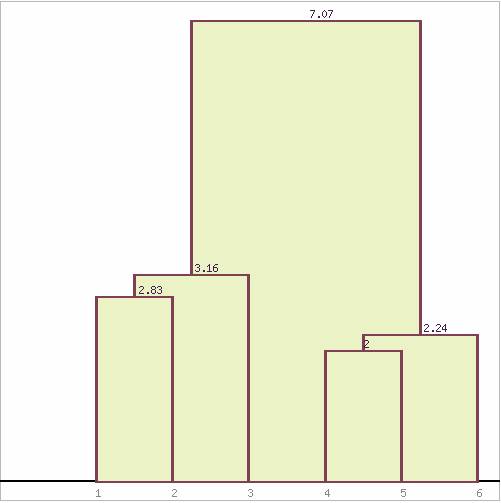
|  |  |  |  |
| --- | --- | --- | --- |
|  | **1,2** | **3** | 4,5,6 |
| **1,2** | 0 | 3.16 | 8.94 |
| **3** | 3.16 | 0 | 7.07 |
| 4,5,6 | 8.94 | 7.07 | 0 |

Продолжаем аналогично. На следующем шаге сливаются в один кластеры 1 и 2. Они наиболее близки.

На последнем шаге объединяются кластеры 1,2 и 3. В результате имеем два кластера – 1,2,3 и 4,5,6.

|  |  |  |
| --- | --- | --- |
|  | 1,2,3 | 4,5,6 |
| 1,2,3 | 0 | 7.07 |
| 4,5,6 | 7.07 | 0 |

Как происходило слияние объектов в кластеры можно представить в виде дендрограммы.



**Задача.** В таблице указаны все попарные расстояния между точками.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** | **F** |
| **A** | 0 | 662 | 877 | 255 | 412 | 996 |
| **B** | 662 | 0 | 295 | 468 | 268 | 400 |
| **C** | 877 | 295 | 0 | 754 | 564 | 138 |
| **D** | 255 | 468 | 754 | 0 | 219 | 869 |
| **E** | 412 | 268 | 564 | 219 | 0 | 669 |
| **F** | 996 | 400 | 138 | 869 | 669 | 0 |

Выполните «вручную» для них агломеративный иерархический алгоритм. Нарисуйте дендограмму кластеров.