

Иерархический кластерный анализ

Введение в кластерный анализ

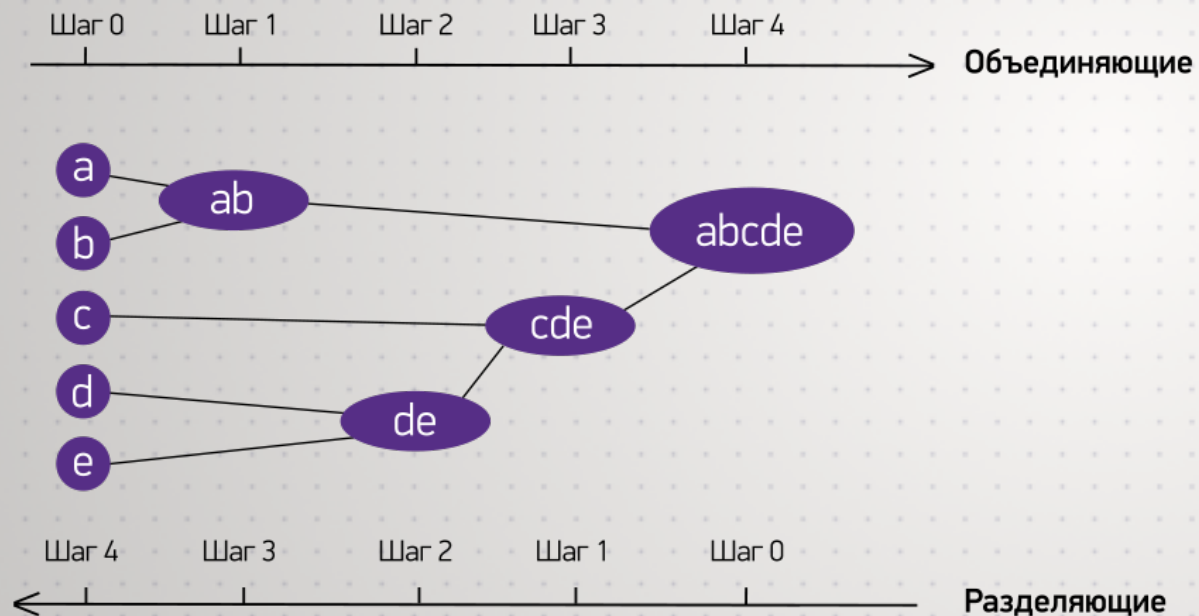
Примеры задач, для которых используется кластерный анализ

- Типологии объектов.
- В социологии: построение моделей стратификации.

Иерархические методы кластерного анализа

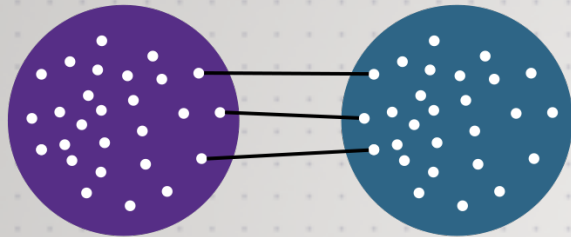
Последовательное, пошаговое объединение (или разделение) объектов.

- Объединяющие, агломеративные (agglomerative) методы.
- Разделяющие, дивизивные (divisive) методы.



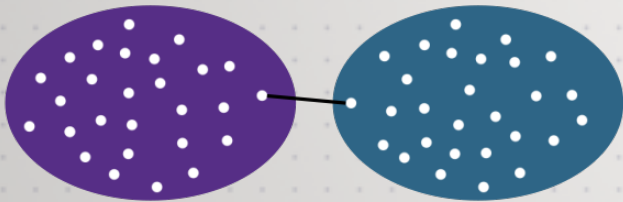
Как происходит построение кластеров?

1. Метод межгрупповой связи.



Хорош для сферических кластеров с равными дисперсиями и объёмами выборок.

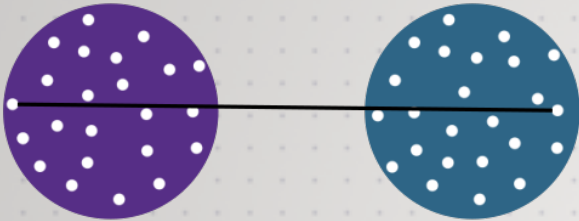
2. Метод ближайшего соседа.



Хорош для растянутых кластеров с разными дисперсиями и объёмами выборок.

Как происходит построение кластеров?

3. Метод удалённого соседа.



Хорош для сферических кластеров с равными дисперсиями и объёмами выборок.

4. Метод центроидной классификации:

Взвешенный центроидный метод.



Невзвешенный центроидный метод.



Как происходит построение кластеров?

5. Метод Уорда

$$V_l = \sum_i \sum_j (x_{ij} - x_{jl})^2$$

l — номер кластера

i — номер объекта ($i = 1, 2, \dots, n_l$)

n_l — количество объектов в l -том кластере

j — номер признака ($j = 1, 2, \dots, k$)

K — количество признаков,
характеризующих каждый объект

Хорош для получения кластеров небольших размеров.

Как определяется сходство?

- Меры расстояния (метрики).
- Меры сходства и ассоциативности