

Метод k-средних

Итерационные методы кластерного анализа

Особенности итерационных методов кластерного анализа

В итерационных методах заранее задается количество кластеров.

Откуда эта информация может быть известна:

1. Априори.
2. Предварительный анализ.

Основная идея метода k-средних

$$\sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \rightarrow \min$$

k — количество кластеров

S_i — кластеры

μ — центры кластеров

Алгоритм кластеризации методом k -средних

1. Выбираем количество кластеров k .
2. Выбираем первоначальные точки центров кластеров m_0 .
3. Считаем расстояние от каждого элемента до центров кластеров.
Распределяем элементы в те кластеры, центры которых оказались ближе к ним.
4. Пересчитываем центры кластеров как центры масс элементов в кластере.
5. Повторяем этап перераспределения элементов.
6. Пересчитываем центры кластеров.
7. Повторяем алгоритм, пока не получим стабильное разбиение.
Либо ориентируемся на заданное количество итераций.

Проблемы метода k-средних

1. Необходимо заранее задать количество кластеров.
2. Результат зависит от первоначального выбора центров кластеров.
3. Не гарантирует сходимость к глобальному минимуму среднеквадратического отклонения от центра кластеров.