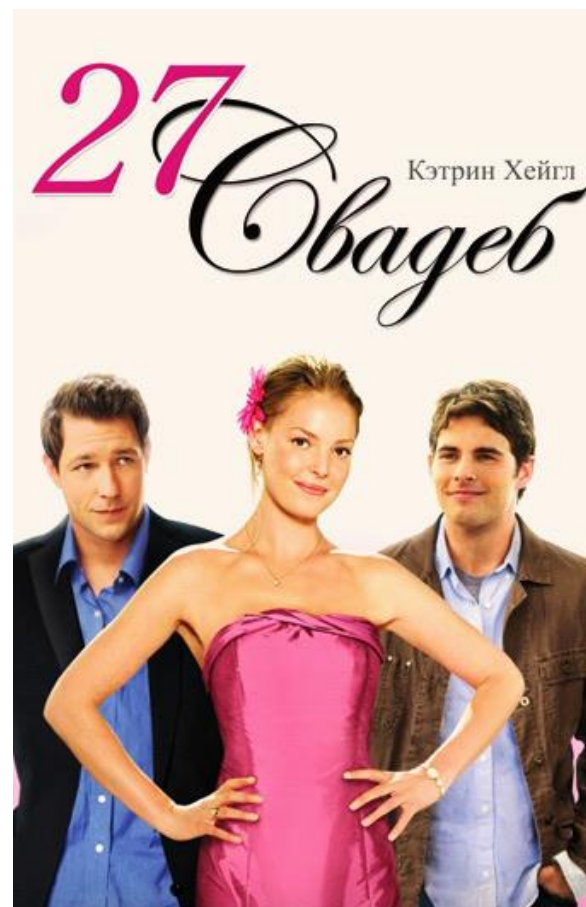


Кластеризация методом k- средних

Лекция 4

Поиск кластеров клиентов



Пример: профили кинозрителей

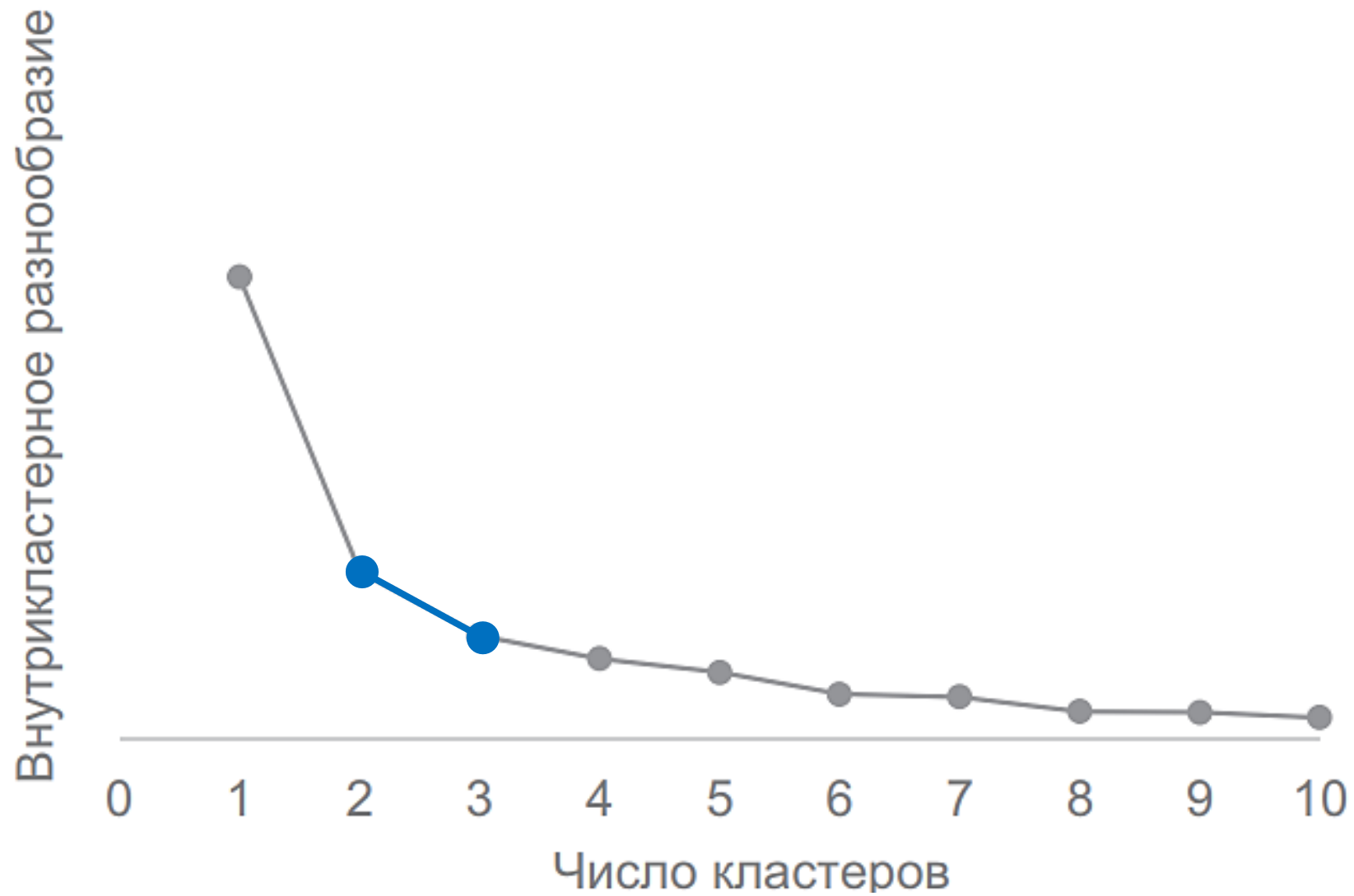
Информация о клиентах:

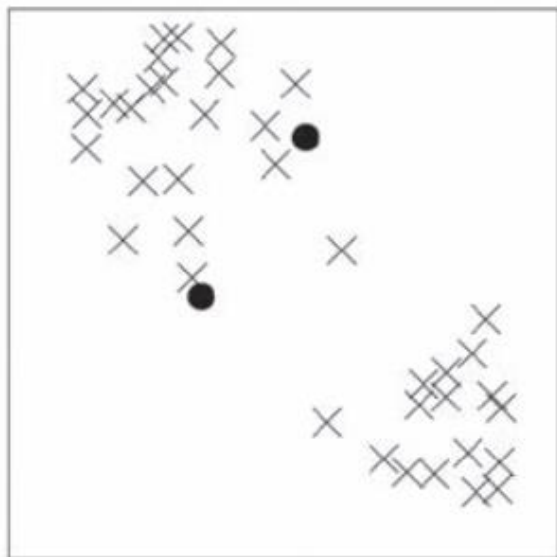
- Доход
 - высокий
 - низкий
- Особенности личности
 - Экстраверсия
 - Добросовестность
 - Эмоциональность
 - Открытость



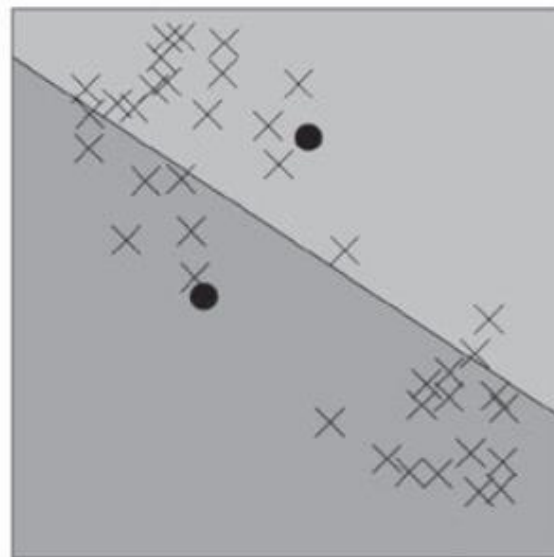
Рис. 1. Персональные профили кинозрителей

Рис. 2. График осыпи показывает изломы, из которых следует, что оптимальное число кластеров от 2 до 3

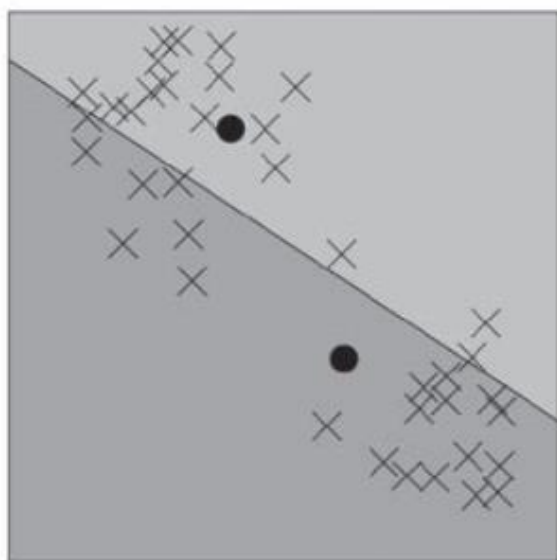




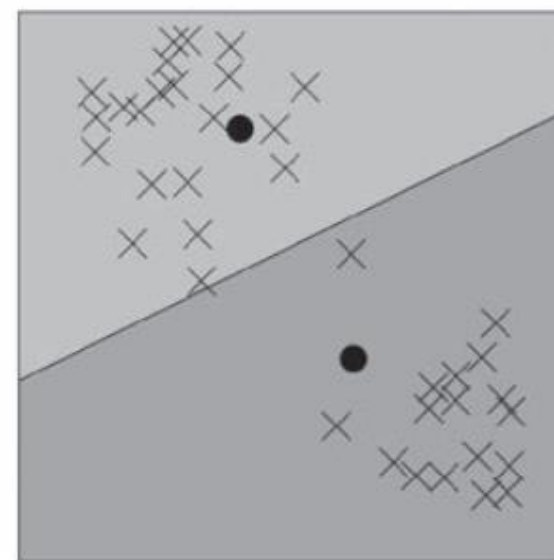
a) шаг 0



b) шаг 1



c) шаг 2



d) шаг 3

Рис. 3. Итеративный процесс кластеризации методом k -средних

Ограничения

- Каждый элемент данных может быть связан только с одним кластером.
- Предполагается, что кластеры сферичны.
- Кластеры предполагаются цельными.

Краткие итоги

- Кластеризация методом k -средних — это способ сгруппировать вместе похожие элементы данных. Число этих кластеров k должно быть указано заранее.
- Для группировки элементов данных сначала каждый из них соотносится с кластером, а потом обновляются позиции кластерных центров. Эти два шага повторяются до тех пор, пока изменения не будут исчерпаны.
- Кластеризация методом k -средних лучше работает для сферичных, непересекающихся кластеров.