

Кластеризация

Импровизация

Кластеризация

- Кластерный анализ (англ. cluster analysis) — многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы.

Кластеризация

- Кластеризация – группировка данных, с той или иной целью, масштабами и инструментами.
- Кластеризация – процесс, когда надо поделить одну кучу ваших данных на несколько, да не наобум, а чтоб с душой и разумом, да для дела (М. В. Ломоносов).

???



BOJGOZ.FU

Методы кластеризации

- По техническому способу сделать кластеризацию.
- По способу отличия одного от другого (по мерам).
- По инструментарию отличия одного от другого.
- Да мало ли чего (добавить по ходу мысли).

Технические способы (методы) кластеризации (основные):

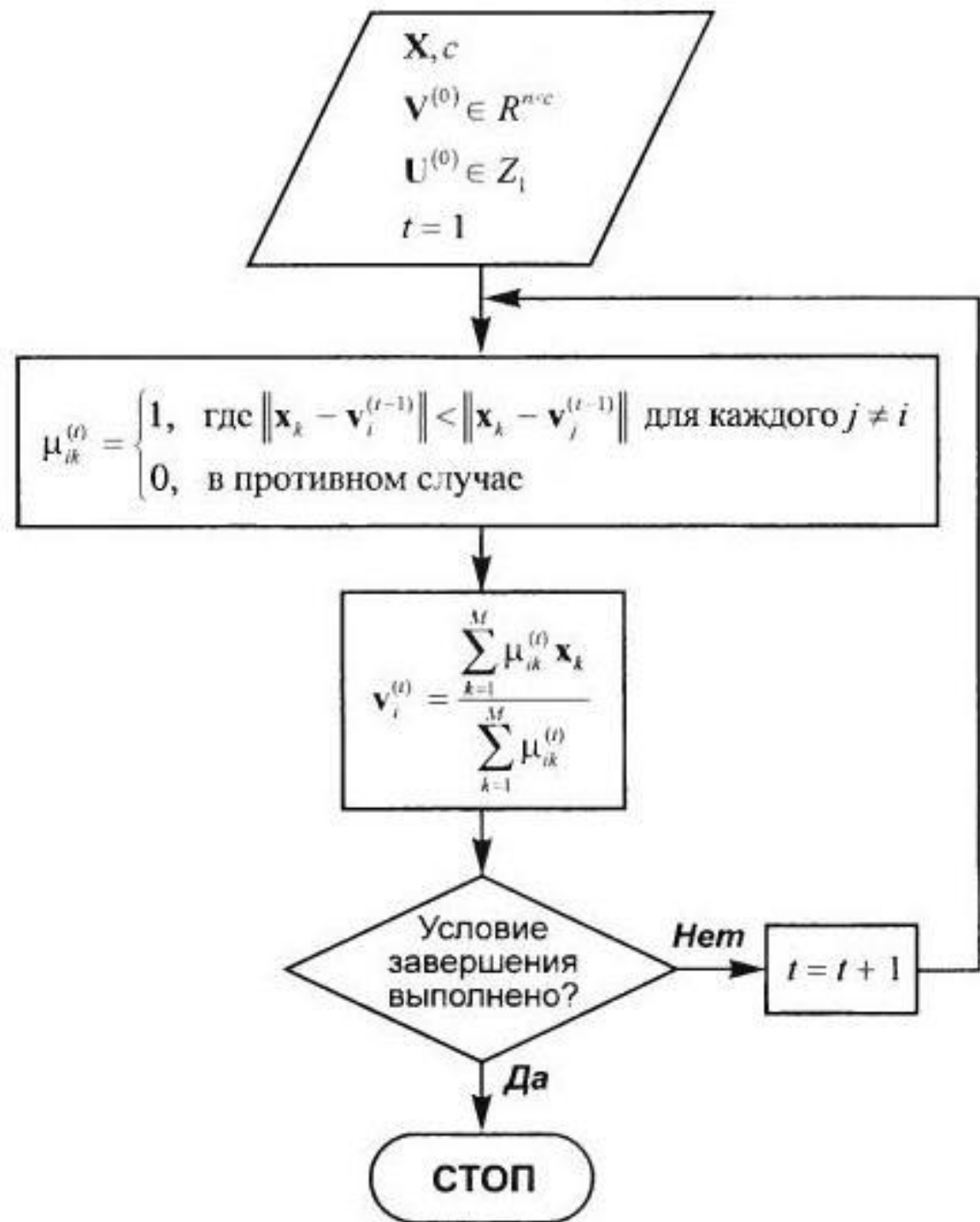
- HCM
- FCM
- PCM
- Густаффсона-Кесселя
- FMLE

Fuzzy logic



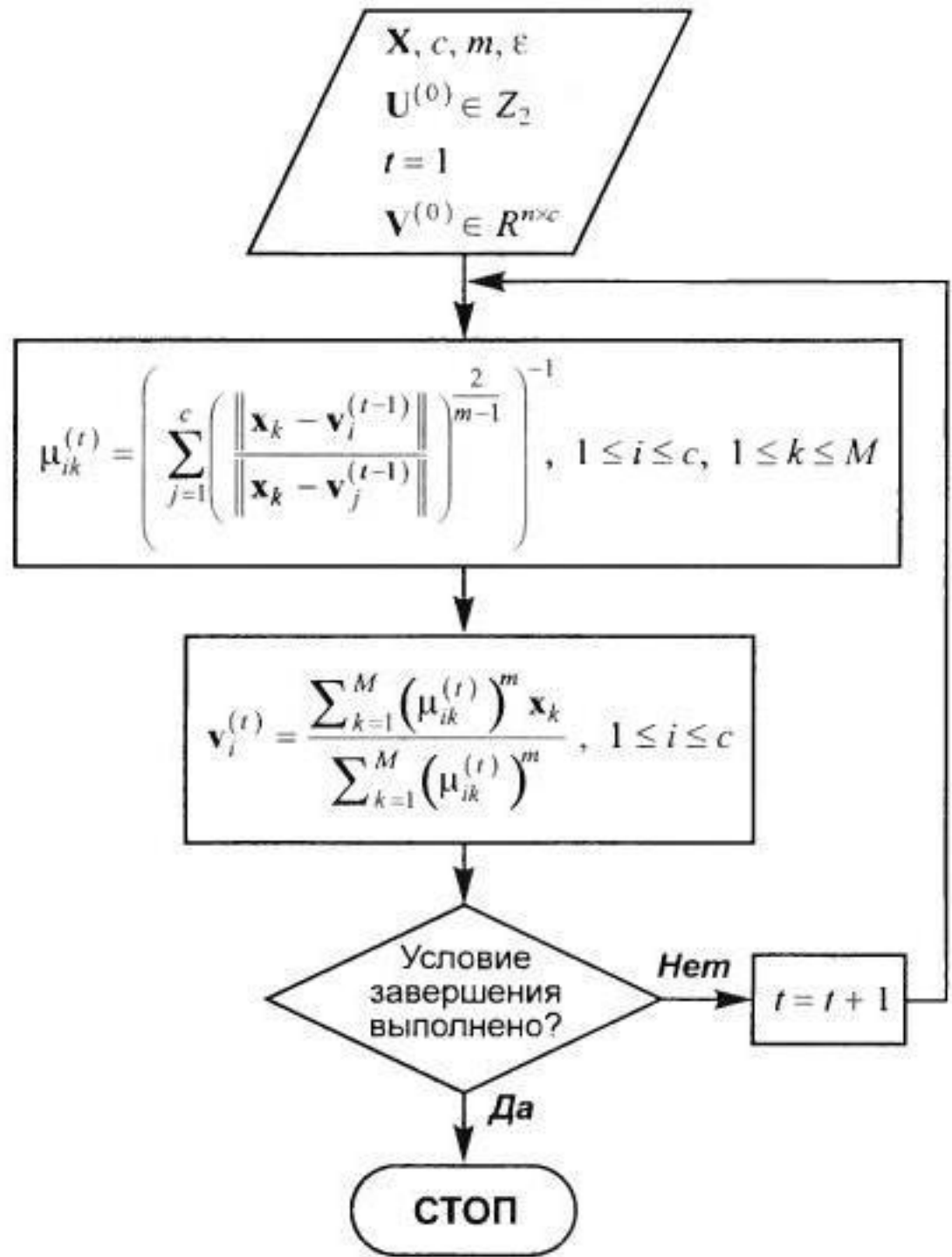
НСМ

- Простота
отвага!



FCSM

- Важно!
Считает сумму принадлежности объекта ко всем кластерам равной 1



PCM

- Уже не всё так просто и весело. Сумма принадлежности не 1

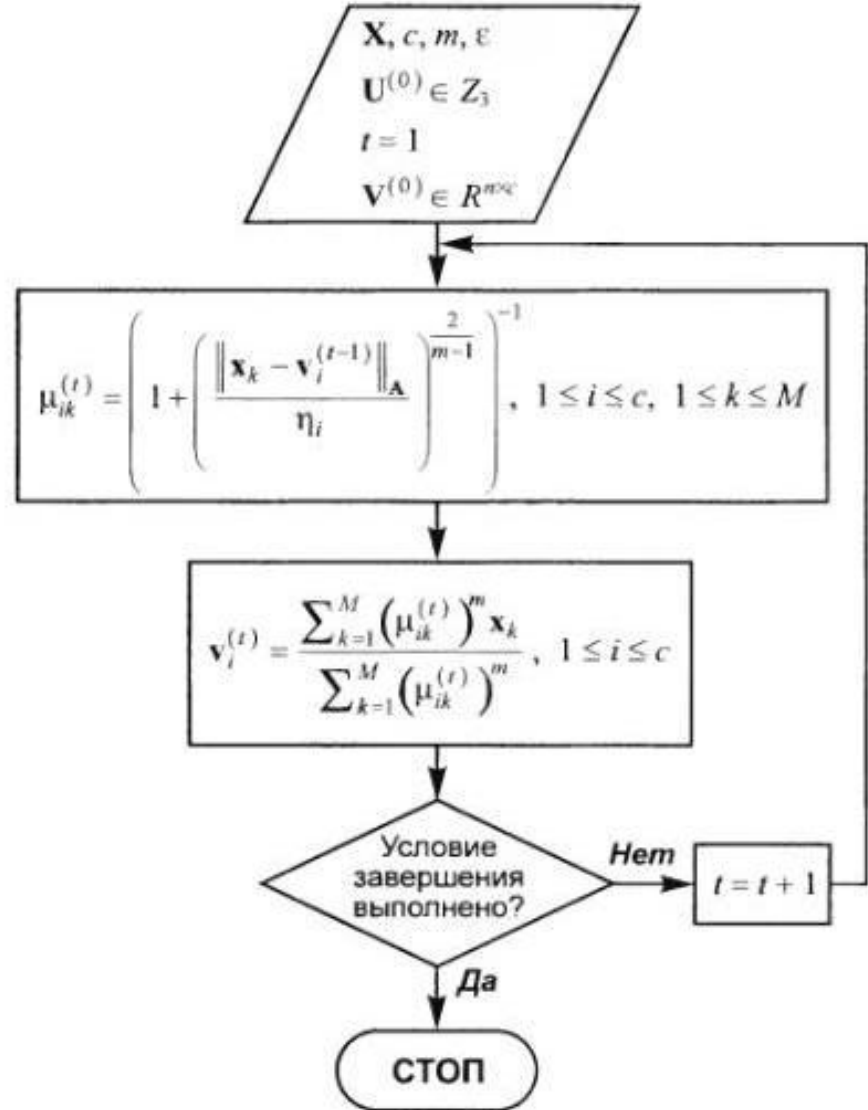
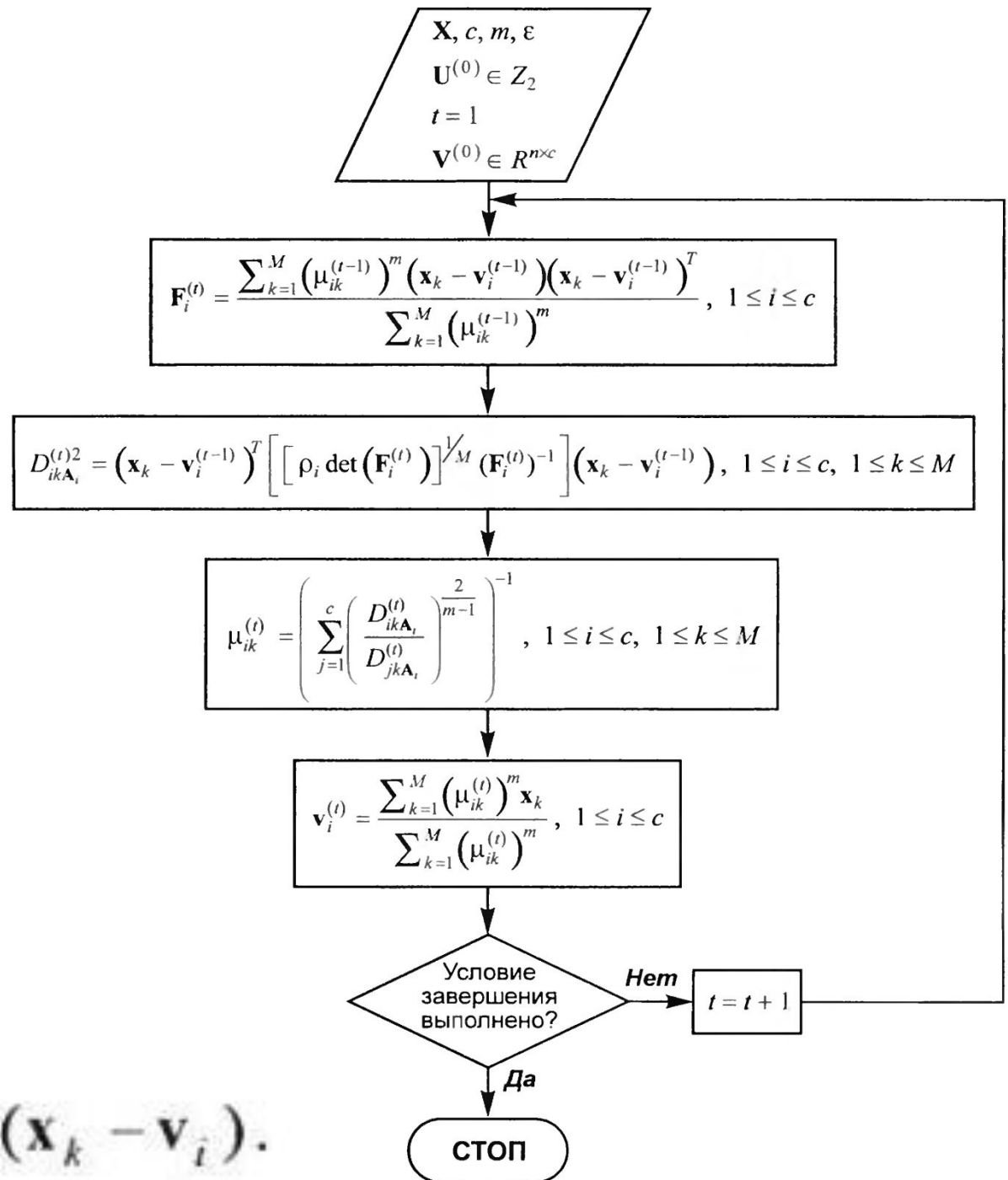


Рис. 8.9. Алгоритм PCM

$$\eta_i = \frac{\sum_{k=1}^M (\mu_{ik})^m D_{ikA}^2}{\sum_{k=1}^M (\mu_{ik})^m}$$

Г-К

- Норма
меняется



$$D_{ik}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{x}_k - \mathbf{v}_i).$$

FMLE



Качество группирования

$$V_1(\mathbf{U}) = \frac{1}{M} \sum_{i=1}^c \sum_{k=1}^M (\mu_{ik})^2.$$

Меры

- Эвклид
- Минковский (Чебышев)
- Махаланобис
- На основе функции выживаемости (модифицированный Махаланобис)
- Расстояние Левенштейна

Методы

- Классические (вы их видели)
- Нейронные сети
- Системы поддержки принятия решений

Практика