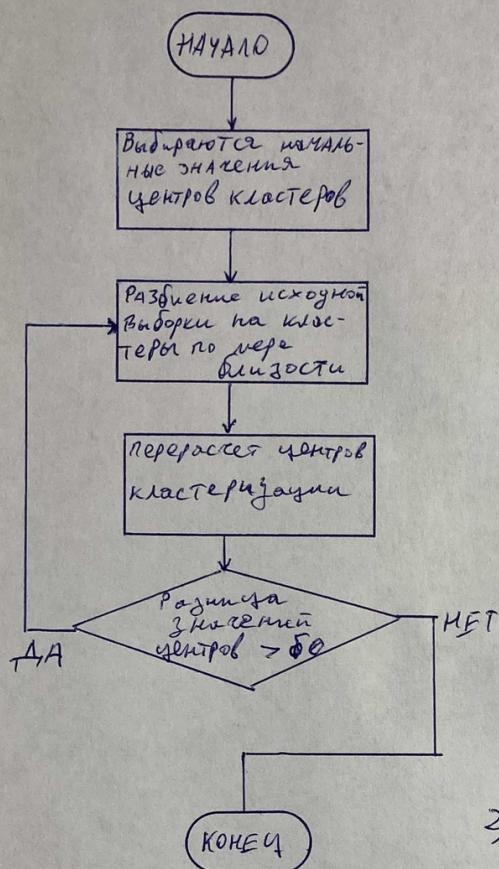


Дуров М.А. Группа 1.2

k-means



1) Выбираются нач. значения m_i , $i = \overline{1, m}$, в качестве которых назначаются либо первые образы из исх. выборки X^N (случайно по-факту), либо m наиболее удаленных друг от друга образов выборки (макс. алгоритм). Это есть начальное приближение, не означает, что оно останется. Это позволяет начать с некоторой выгодной инициализации.

2) Проводится разбиение $Q_m = \{x_1^{m_1}, \dots, x_{m_m}^{m_m}\}$

исходной выборки на кластеры (по мере близости - берем кажд. образ из обусловленной выборки, находим расст. до центра и относим в тот кластер, где расст. это будет min.) в ходе которого принадлежность кажд. образа опреу. на основе поиска минимума среди расстояний до установленных ранее центров m_i , $i = \overline{1, m}$. Рассчит. начальное значение внутриклассового разброса $E_w(X^N, Q_m)$

3) Проводится перерасчет центров с учетом рез-тов выполненной в п.2 кластеризации

$$m'_i = \frac{1}{N_i} \sum_{k=1}^{N_i} x^{(k,i)}, \quad i = \overline{1, m} \quad \text{и соотв. перерасчет}$$

ф-ии $E_w(X^N, Q_m)$.

4) Если разность значений центров на соседних шагах $\delta_{\max} = \max |m_i - m'_i| > \delta_0$, то осуществл.

присвоение $m_i = m'_i, i = \overline{1, m}$ и переход к выполнению след. шага в п. 2. Число - остановится.