Маликов С.А. 2.2 группа

Алгоритмы кластеризации по методу агломеративной иерархической группировки

Мат. описание:

шаг 1: пусть имеется:

- не помеченная, смешанная выборка образов $X^N = \{x^{(1)}, \dots, x^{(N)}\}, x^{(*)} \in R^n$
- M - число классов
- $d(x^{(i)}, x^{(j)})$ - функция расстояния.

шаг 2: устанавливаем значение $\tilde{M}$, $\tilde{M} = N$, N - количество образов в выборке образов.

делаем начальную кластеризацию Q_N^1, $Q_N^1 = \{\overset{1}{x}, \dots, \overset{N}{x}\}, \overset{j}{x} = x^{(j)}$.

следующие шаги выполняем в цикле.

шаг 3:

Для Q^k_{N-k+1} и $\tilde{M} = N - k + 1$ находим ближайшую пару групп с использованием $d(x^{(i)}, x^{(j)})$, которая имеет $\min\limits_{i,j} D(x^{N_i}, x^{N_j})$.

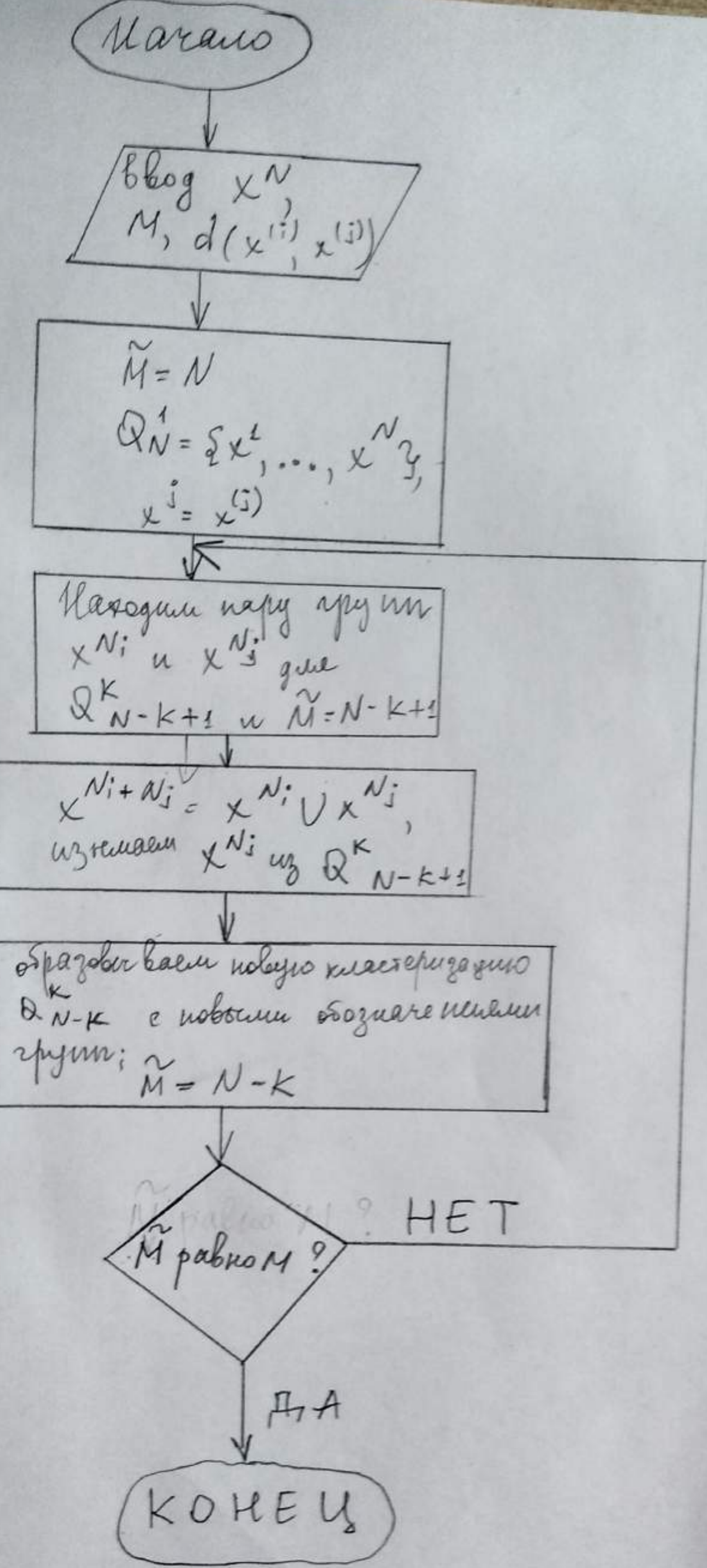
шаг 4: производим слияние найденных групп $x^{N_i + N_j} = x^{N_i} \cup x^{N_j}$.

изымаем из Q^k_{N-k+1} элемент x^{N_j}.

шаг 5: образовываем новую кластеризацию Q^k_{N-k} с новыми обозначениями групп и присвоением $\tilde{M} = N - K$

шаг 6: сравниваем значение $\tilde{M}$ и M, и если $\tilde{M} = M$, то ~~алгоритм~~ цикл завершается, иначе возвращаемся к шагу 3.

Замечание: если в шаге 1 значение числа классов не задается, то в шаге 6 значение $\tilde{M}$ сравнивают с единицей.


Начало
ввод X^N, M, $d(x^{(i)}, x^{(j)})$
$\tilde{M} = N$
$Q_N^1 = \{X^1, \dots, X^N\}$, $x^j = x^{(j)}$
Находим пару групп X^{N_i} и X^{N_j} для Q^K_{N-K+1} и $\tilde{M} = N-K+1$
$X^{N_i+N_j} = X^{N_i} \cup X^{N_j}$, изымаем X^{N_j} из Q^K_{N-K+1}
образовываем новую кластеризацию Q^K_{N-K} с новыми обозначениями групп; $\tilde{M} = N-K$
$\tilde{M}$ равно M?
НЕТ
ДА
КОНЕЦ