Peter the Great
Saint-Petersburg Polytechnic University

«Применение нечеткой кластеризации к дендритным шипикам»

Исполнитель: Лахнева М. Ю.

01.04.23

POLYTECH

Алгоритм нечетких C-средних (FCM)

Введем следующие обозначения:

$d \in \mathbb{N}$ – размерность пространства векторов данных;

$l \in \mathbb{N}: 1 \leq l \leq d$ – номер координаты вектора;

$n \in \mathbb{N}$ – мощность обучающей выборки;

$\mathrm{X} \subset \mathbb{R}^d$ – обучающая выборка векторов данных;

$i \in \mathbb{N}: 1 \leq i \leq n$ – номер вектора обучающей выборки;

$x_i \in \mathrm{X}$ – i–й вектор выборки;

$k \in \mathbb{N}$ – количество кластеров;

$j \in \mathbb{N}: 1 \leq j \leq k$ – номер кластера;

$C \subset \mathbb{R}^{k \times d}$ – матрица, которая содержит центры кластеров;

$c_i \in \mathbb{R}^d$ – центр кластера j, вектор размерности d;

$x_{il}, c_{il} \in \mathbb{R}$ – l–е координаты векторов x_i и c_iсоответственно;

$U \subset \mathbb{R}^{n \times k}$ – матрица степеней принадлежности, где $u_{il} \in \mathbb{R}$:

$0 \leq u_{ij} \leq 1$ – степень принадлежности вектора x_i кластеру j;

$\rho(x_i, c_i)$ – функция расстояния, определяющая степень принадлежности вектора x_i кластеру j;

$m \in \mathbb{R}: m > 1$ – степень нечеткости целевой функции;

J_{FCM} – целевая функция алгоритма FCM.

Алгоритм нечетких С-средних (FCM)

Алгоритм основан на минимизации целевой функции J_{FCM} :

$$J_{FCM}(X,k,m) = \sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij}^{m}\rho^{2}\left(x_i, c_j\right) \quad (1)$$

$$u_{ij} = \sum_{t=1}^{k}\left(\frac{\rho\left(x_i,c_j\right)}{\rho\left(x_i,c_t\right)}\right)^{\frac{2}{1-m}} \quad (2)$$

$$\forall j,l \quad c_{jl} = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_{il}}{\sum_{i=1}^{n} u_{ij}^{m}} \quad (3)$$

Условие завершения алгоритма:

$$\max_{ij}\left\{\left|u_{ij}^{(s+1)} - u_{ij}^{(s)}\right|\right\} < \varepsilon \quad (4)$$

s – номер итерации

$\varepsilon \in (0,1) \subset \mathbb{R}$ – критерий останова

Алгоритм нечетких C-средних (FCM)

Вход: X, m, ε, k.

Выход: U.

Шаг 1. $s := 0$.

Шаг 2. $U^{(0)} := (u_{ij})$.

Шаг 3. (вычисление новых координат центроидов)

Вычислить $C^{(s)} := (c_j)$, используя формулу (3), где $u_{ij} \in U^{(s)}$.

Шаг 4. (обновление значений матриц)

Вычислить $U^{(s)}$ и $U^{(s+1)}$ по формуле (2).

Шаг 5. $s := s + 1$.

Шаг 6. Если условие (4) не выполняется, то перейти на шаг 3.

Шаг 7. Стоп.

Алгоритм нечетких С-средних (FCM)

Пример:

Дано: {(1; 3), (2; 5), (4; 8), (7; 9)}

- Шаг 1: s=0;
- Шаг 2: Инициализируем таблицу принадлежности случайными значениями.

Cluster	(1; 3)	(2; 5)	(4; 8)	(7; 9)
1	0.8	0.7	0.2	0.1
2	0.2	0.3	0.8	0.9

- Шаг 3: вычисление координат центроидов: $c_{jl} = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_{il}}{\sum_{i=1}^{n} u_{ij}^{m}}$

$$c_{11} = \frac{(0.8^2 * 1 + 0.7^2 * 2 + 0.2^2 * 4 + 0.1^2 * 7)}{(0.8^2 + 0.7^2 + 0.2^2 + 0.1^2)} = 1.568 \ldots$$

POLYTECH

Алгоритм нечетких С-средних (FCM)

- Шаг 3: вычисление координат центроидов:

c_{11}= 1.568

c_{12}= 4.051

$$(1.568;\ 4.051)$$

c_{21}= 5.35

c_{22}= 8.215

$$(5.35;\ 8.215)$$

- Шаг 4: обновление матриц принадлежности:

Рассчитаем расстояние между точками и центром тяжести с помощью метрики – евклидово расстояние:
$\{(1;3),(2;5),(4;8),(7;9)\}$

$$\rho(x_1, c_1) = \sqrt{(1-1.568)^2 + (3-4.051)^2} = 1.2; \qquad \rho(x_3, c_1) = 4.63$$
$$\rho(x_1, c_2) = 6.79; \qquad \rho(x_3, c_2) = 1.36$$
$$\rho(x_2, c_1) = 1.04; \qquad \rho(x_4, c_1) = 7.34$$
$$\rho(x_2, c_2) = 4.64; \qquad \rho(x_4, c_2) = 1.82$$

Алгоритм нечетких C-средних (FCM)

- Координат центроидов:

первый - (1.568; 4.051)

второй - (5.35; 8.215)

- Расстояния между точками и центром тяжести:

$$\rho(x_1, c_1) = 1.2; \quad \rho(x_2, c_1) = 1.04; \ \rho(x_3, c_1) = 4.63; \ \rho(x_4, c_1) = 7.34$$
$$\rho(x_1, c_2) = 6.79; \ \rho(x_2, c_2) = 4.64; \ \rho(x_3, c_2) = 1.36; \ \rho(x_4, c_2) = 1.82$$

Cluster	(1; 3)	(2; 5)	(4; 8)	(7; 9)
1	0.8	0.7	0.2	0.1
2	0.2	0.3	0.8	0.9
центроид	первый	первый	второй	второй

Обновляем матрицу принадлежности:

$$u_{ij} = \sum_{t=1}^{k} \left(\frac{\rho(x_i, c_j)^2}{\rho(x_i, c_t)^2} \right)^{\frac{1}{1-m}}$$

Алгоритм нечетких С-средних (FCM)

$$u_{11} = \left(\frac{1.2^2}{1.2^2} + \frac{1.2^2}{6.79^2}\right)^{\left(\frac{1}{1-2}\right)} = 0.97; \qquad u_{31} = 0.08$$

$$u_{12} = 0.03; \qquad u_{32} = 0.92$$

$$u_{21} = 0.95; \qquad u_{41} = 0.06$$

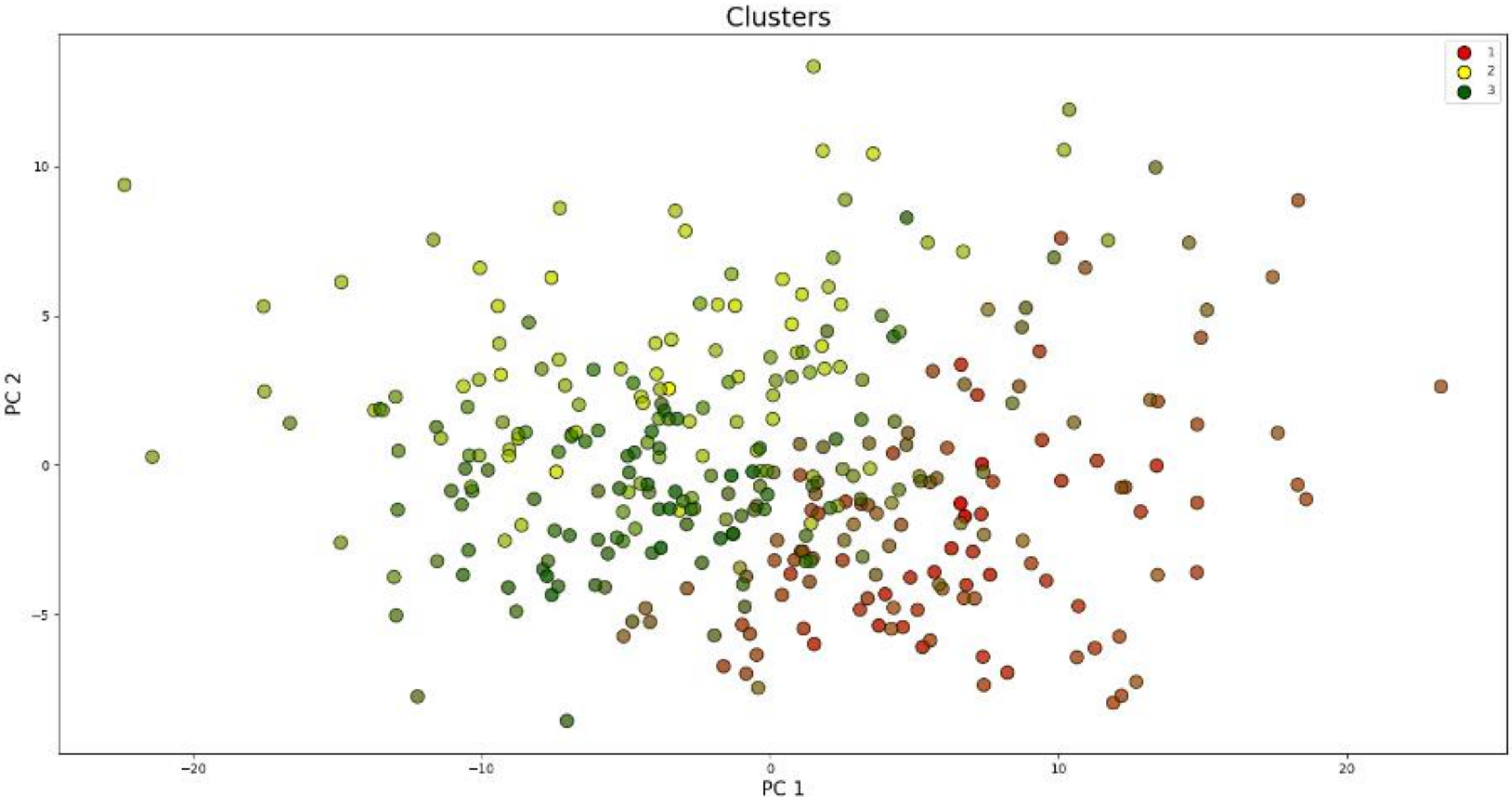
$$u_{22} = 0.05; \qquad u_{42} = 0.94$$

	Cluster	(1; 3)	(2; 5)	(4; 8)	(7; 9)
s=0	1	0.8	0.7	0.2	0.1
	2	0.2	0.3	0.8	0.9
s=1	1	0.97	0.95	0.08	0.06
	2	0.03	0.05	0.92	0.94

- Шаг 5: s=1;
- Шаг 6: проверяем условие:

$$\max_{ij}\left\{\left|u_{ij}^{(s+1)} - u_{ij}^{(s)}\right|\right\} < \varepsilon$$

FCM



POLYTECH