

Результаты выполнения алгоритма k-means

Набор данных

Набор данных состоит из 16 точек в 4-мерном пространстве (X_1, X_2, X_3, X_4) :

Точка	X_1	X_2	X_3	X_4
x_0	-5	-12	12	10
x_1	-4	-5	9	-15
x_2	-3	-7	11	7
x_3	-4	-11	-5	13
x_4	-20	-18	0	19
x_5	-11	3	17	3
x_6	18	-10	1	11
x_7	2	-5	9	8
x_8	17	18	-20	2
x_9	12	0	1	-1
x_{10}	-13	-4	-7	-17
x_{11}	-20	0	-10	8
x_{12}	10	-18	12	-3
x_{13}	3	-13	-5	8
x_{14}	-17	1	-4	-5
x_{15}	-3	-5	16	12

Таблица 1: Набор данных

Параметры

- Количество кластеров: $k = 3$
- Инициализация: Forgy (случайный выбор точек x_0, x_6, x_8)
- Начальные центроиды:
 - $\mu_1 = [-5, -12, 12, 10]$
 - $\mu_2 = [18, -10, 1, 11]$
 - $\mu_3 = [17, 18, -20, 2]$

Формулы

1. Евклидово расстояние между точкой $x_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}]$ и центроидом $\mu_j = [\mu_{j1}, \mu_{j2}, \mu_{j3}, \mu_{j4}]$:

$$d(x_i, \mu_j) = \sqrt{\sum_{m=1}^4 (x_{im} - \mu_{jm})^2}$$

2. Обновление центроида для кластера S_j , содержащего $|S_j|$ точек:

$$\mu_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$$

где $\mu_{jm} = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_{im}$, $m = 1, 2, 3, 4$.

Итерация 1

Распределение точек

Для каждой точки x_i вычисляется евклидово расстояние до центроидов μ_1, μ_2, μ_3 , и точка присваивается кластеру с минимальным расстоянием.

Пример для $x_0 = [-5, -12, 12, 10]$:

До $\mu_1 = [-5, -12, 12, 10]$:

$$d(x_0, \mu_1) = \sqrt{(-5 - (-5))^2 + (-12 - (-12))^2 + (12 - 12)^2 + (10 - 10)^2} = \sqrt{0} = 0$$

До $\mu_2 = [18, -10, 1, 11]$:

$$\begin{aligned} d(x_0, \mu_2) &= \sqrt{(-5 - 18)^2 + (-12 - (-10))^2 + (12 - 1)^2 + (10 - 11)^2} \\ &= \sqrt{(-23)^2 + (-2)^2 + 11^2 + (-1)^2} = \sqrt{529 + 4 + 121 + 1} = \sqrt{655} \approx 25.59 \end{aligned}$$

До $\mu_3 = [17, 18, -20, 2]$:

$$\begin{aligned} d(x_0, \mu_3) &= \sqrt{(-5 - 17)^2 + (-12 - 18)^2 + (12 - (-20))^2 + (10 - 2)^2} \\ &= \sqrt{(-22)^2 + (-30)^2 + 32^2 + 8^2} = \sqrt{484 + 900 + 1024 + 64} = \sqrt{2472} \approx 49.72 \end{aligned}$$

Метка: $cluster_labels[0] = 1$.

Итоговые метки: $[1, 1, 1, 1, 1, 1, 2, 1, 3, 2, 1, 1, 2, 1, 1, 1]$.

Кластеры

- Кластер 1 (S_1): точки 0, 1, 2, 3, 4, 5, 7, 10, 11, 13, 14, 15 ($|S_1| = 12$)
- Кластер 2 (S_2): точки 6, 9, 12 ($|S_2| = 3$)
- Кластер 3 (S_3): точка 8 ($|S_3| = 1$)

Обновление центроидов

Для S_1 :

$$\mu_1 = \frac{1}{12} \sum_{i \in \{0,1,2,3,4,5,7,10,11,13,14,15\}} x_i$$

- $X_1: -5 + (-4) + (-3) + (-4) + (-20) + (-11) + 2 + (-13) + (-20) + 3 + (-17) + (-3) = -95$, $\mu_{1,1} = \frac{-95}{12} \approx -7.9167$
- $X_2: -12 + (-5) + (-7) + (-11) + (-18) + 3 + (-5) + (-4) + 0 + (-13) + 1 + (-5) = -76$, $\mu_{1,2} = \frac{-76}{12} \approx -6.3333$
- $X_3: 12 + 9 + 11 + (-5) + 0 + 17 + 9 + (-7) + (-10) + (-5) + (-4) + 16 = 43$, $\mu_{1,3} = \frac{43}{12} \approx 3.5833$
- $X_4: 10 + (-15) + 7 + 13 + 19 + 3 + 8 + (-17) + 8 + 8 + (-5) + 12 = 51$, $\mu_{1,4} = \frac{51}{12} = 4.25$

$$\mu_1 \approx [-7.9167, -6.3333, 3.5833, 4.25]$$

Для S_2 :

$$\mu_2 = \frac{1}{3} \sum_{i \in \{6,9,12\}} x_i$$

- $X_1: 18 + 12 + 10 = 40, \mu_{2,1} = \frac{40}{3} \approx 13.3333$
- $X_2: -10 + 0 + (-18) = -28, \mu_{2,2} = \frac{-28}{3} \approx -9.3333$
- $X_3: 1 + 1 + 12 = 14, \mu_{2,3} = \frac{14}{3} \approx 4.6667$
- $X_4: 11 + (-1) + (-3) = 7, \mu_{2,4} = \frac{7}{3} \approx 2.3333$

$$\mu_2 \approx [13.3333, -9.3333, 4.6667, 2.3333]$$

Для $S_3: \mu_3 = [17, 18, -20, 2]$.

Сходимость

$\mu_1 \neq [-7.9167, -6.3333, 3.5833, 4.25]$, поэтому *converges = False*.

Новые центроиды:

$$\mu_1 \approx [-7.9167, -6.3333, 3.5833, 4.25], \quad \mu_2 \approx [13.3333, -9.3333, 4.6667, 2.3333], \quad \mu_3 = [17, 18, -20, 2]$$

Итерация 2

Распределение точек

Пример для $x_6 = [18, -10, 1, 11]$:

До $\mu_1 = [-7.9167, -6.3333, 3.5833, 4.25]$:

$$\begin{aligned} d(x_6, \mu_1) &= \sqrt{(18 - (-7.9167))^2 + (-10 - (-6.3333))^2 + (1 - 3.5833)^2 + (11 - 4.25)^2} \\ &= \sqrt{(25.9167)^2 + (-3.6667)^2 + (-2.5833)^2 + (6.75)^2} \\ &= \sqrt{671.67 + 13.44 + 6.67 + 45.56} = \sqrt{737.34} \approx 27.15 \end{aligned}$$

До $\mu_2 = [18, -10, 1, 11]$:

$$\begin{aligned} d(x_6, \mu_2) &= \sqrt{(18 - 13.3333)^2 + (-10 - (-9.3333))^2 + (1 - 4.6667)^2 + (11 - 2.3333)^2} \\ &= \sqrt{(4.6667)^2 + (-0.6667)^2 + (-3.6667)^2 + (8.6667)^2} \\ &= \sqrt{21.78 + 0.44 + 13.44 + 75.11} = \sqrt{110.77} \approx 10.52 \end{aligned}$$

До $\mu_3 = [17, 18, -20, 2]$:

$$\begin{aligned} d(x_6, \mu_3) &= \sqrt{(18 - 17)^2 + (-10 - 18)^2 + (1 - (-20))^2 + (11 - 2)^2} \\ &= \sqrt{1^2 + (-28)^2 + 21^2 + 9^2} = \sqrt{1 + 784 + 441 + 81} = \sqrt{1307} \approx 36.15 \end{aligned}$$

Метка: $cluster_labels[6] = 2$.

Метки: $[1, 1, 1, 1, 1, 1, 2, 1, 3, 2, 1, 1, 2, 1, 1, 1]$.

Кластеры

Не изменились.

Обновление центроидов

Центроиды идентичны:

$$\mu_1 \approx [-7.9167, -6.3333, 3.5833, 4.25], \quad \mu_2 \approx [13.3333, -9.3333, 4.6667, 2.3333], \quad \mu_3 = [17, 18, -20, 2]$$

Сходимость

converges = True.

Финальные результаты

- Метки кластеров: $[1, 1, 1, 1, 1, 1, 2, 1, 3, 2, 1, 1, 2, 1, 1, 1]$

- Центроиды:

$$\mu_1 \approx [-7.9167, -6.3333, 3.5833, 4.25]$$

$$\mu_2 \approx [13.3333, -9.3333, 4.6667, 2.3333]$$

$$\mu_3 = [17, 18, -20, 2]$$

Примечания

Результаты зависят от начальной инициализации. Для другого выбора центроидов кластеры могут отличаться.