**Пример решения задачи кластеризации методом k-средних**

[**http://axd.semestr.ru/upr/k-means.php**](http://axd.semestr.ru/upr/k-means.php)

По таблице с данными провести классификацию объектов на **три кластера** методом К-средних. Провести максимальное число итераций. Эталонные точки и порядок появления точек задать самостоятельно.

Отобразить на плоскости полученный вариант классификации.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Объекты | A | B | C | D | E | F |
| признак-X | –2 | –2 | –3 | 4 | 3 | 0 |
| признак-Y | 0 | –1 | 0 | 0 | 0 | 0 |

**Решение** проводим с помощью [калькулятора](http://axd.semestr.ru/upr/average.php).

**Итерация №1**

Эталонные точки - центроиды кластеров:

e1 = A = (-2; 0)

e2 = B = (-2; -1)

e3 = C = (-3; 0)

Сравниваем расстояние от точки D до эталонных точек.

D(D,e1) = **√** (4 - (-2))2 +(0 – 0)2 = **√** 36 = 6

D(D,e2) = **√** (4 - (-2))2 +(0 – (-1)2 = **√** (64 + 1) = **√** 65

D(D,e3) = **√** (4 - (-3))2 +(0 – 0)2 = **√** 49 = 7

Минимальным является расстояние d(D;e1)

Сравниваем расстояние от точки E до эталонных точек.

D(E,e1) = **√** (3 - (-2))2 +(0 – 0)2 = **√** 25 = 5

D(E,e2) = **√** (3 - (-2))2 +(0 – (-1)2 = **√** (25 + 1) = **√** 26

D(E,e3) = **√** (3 - (-3))2 +(0 – 0)2 = **√** 36 = 6

Минимальным является расстояние d(E,e1)

Сравниваем расстояние от точки F до эталонных точек.

D(F,e1) = **√** (0 - (-2))2 +(0 – 0)2 = **√** 16 = 4

D(F,e2) = **√** (0 - (-2))2 +(0 – (-1)2 = **√** (4 + 1) = **√** 5

D(F,e3) = **√** (0 - (-3))2 +(0 – 0)2 = **√** 9 = 3

Минимальным является расстояние d(F;e2)

Выписываем в таблице принадлежность объектов кластерам

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Объект | A | B | C | D | E | F |
| признак-X | –2 | –2 | –3 | 4 | 3 | 0 |
| признак-Y | 0 | –1 | 0 | 0 | 0 | 0 |
| **Номер кластера** | **1** | **2** | **3** | **1** | **1** | **2** |

|  |  |  |
| --- | --- | --- |
| e1 | e2 | e3 |
| A,D,E | B,F | C |

Пересчитываем значения для эталонных точек.

e1: (-2+4+3)/3 = 5/3; (0+0+0)/3 = 0; e1 =(5/3; 0)

e2:; (-2+0)/2=-1: (-1+0)=-1/2 e2 =(-1; -1/2)

e3: -3; 0 e3 =(-3; 0)

**Итерация №2**

Эталонные точки - центроиды кластеров:

e1 = (5/3; 0)

e2 = (-1; -1/2)

e3 = (-3; 0)

Сравниваем расстояние от точки A(-2;0) до эталонных точек.

D(A,e1) = **√** (-2 – 5/3)2 +(0 – 0)2 = **√** (11/3) 2 = **√** 121/9

D(A,e2) = **√** (-2 – (-1)2 +(0 – (-0,5)2 = **√** (1 +0,25) = **√** 5/4

D(A,e3) = **√** (-2 – (-3/2)2 +(0 – 0)2 = **√** ¼=1/2

Минимальным является расстояние d(A,e3)

Сравниваем расстояние от точки B(-2; -1) до эталонных точек.

D(B,e1) = **√** (-2 - 5/3)2 +(-1 – 0)2 = **√** 121/9 +1 =

D(B,e2) = **√** (-2 – (-1)2 +(-1 – (-0,5))2 = **√** (1 + 0,25) =

D(B,e3) = **√** (-2 - (-3/2))2 +(-1 – 0)2 = **√** 1/4+1 =

Минимальным является расстояние d(B,e2) и d(B,e3)

Сравниваем расстояние от точки C(-3; 0) до эталонных точек.

D(C,e1) = **√** (-3 – 5/3)2 +(0 – 0)2 = **√** (14/3) 2 = **√** 156/9

D(C,e2) = **√** (-3 - (-1))2 +(0 – (-0,5))2 = **√** (4 + 0,25) = **√** 4,25

D(C,e3) = **√** (-3 - (-3/2))2 +(0 – 0)2 = **√** 9/4 = 3/2

Минимальным является расстояние d(C,e1)

Сравниваем расстояние от точки D(4; 0) до эталонных точек.

D(D,e1) = **√** (4 – 5/3)2 +(0 – 0)2 = **√** (7/3) 2 = **√** 49/9=**√** 5+4/9

D(D,e2) = **√** (4 - (-1))2 +(0 – (-0,5))2 = **√** (25 + 0,25) = **√** 25,25

D(D,e3) = **√** (4 - (-3/2))2 +(0 – 0)2 = **√** (5,5) 2

Минимальным является расстояние d(D,e2)

Сравниваем расстояние от точки E(3; 0) до эталонных точек.

D(E,e1) = **√** (3 – 5/3)2 +(0 – 0)2 = **√** 4/9 = 2/3

D(E,e2) = **√** (3 - (-1))2 +(0 – (-0,5))2 = **√** (16 + 0,25) = **√** 16,25

D(E,e3) = **√** (3 - (-3/2))2 +(0 – 0)2 = 4,5

Минимальным является расстояние d(E,e1)

Сравниваем расстояние от точки F(0; 0) до эталонных точек.

D(F,e1) = **√** (0 – 3/5)2 +(0 – 0)2 = **√** 9/25 = 1

D(F,e2) = **√** (0 - (-1))2 +(0 – (-0,5))2 = **√** (1 + 0,25) = **√** 1,25

D(F,e3) = **√** (0 - (-3/2))2 +(0 – 0)2 = 3/2

Минимальным является расстояние d(F,e1)

Выписываем в таблице принадлежность объектов кластерам

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Объект | A | B | C | D | E | F |
| признак-X | –2 | –2 | –3 | 4 | 3 | 0 |
| признак-Y | 0 | –1 | 0 | 0 | 0 | 0 |
| Номер кластера | 3 | 2 | 1 | 2 | 1 | 1 |

|  |  |  |
| --- | --- | --- |
| e1 | e2 | e3 |
| С, E, F | B, D | A |

Пересчитываем значения для эталонных точек.

e1: (-3+3+0)/3 = 0; (0+0+0)/3 = 0; e1 =(0;0)

e2: (-2+4)/2 = 1; (-1+0)/2 = -1/2; e2 =(1; -1/2)

e3: -2; 0; e3 =(-2;0)

Границы кластеров изменились, продолжаем процесс разбиения.

**Итерация №3**

Эталонные точки - центроиды кластеров:

e1 = (0; 0)

e2 = (1; -1/2)

e3 = (-2; 0)

Сравниваем расстояние от точки A(-2;0) до эталонных точек.

D(A,e1) = **√** (-2 –0)2 +(0 – 0)2 = **√** 4 = **2**

D(A,e2) = **√** (-2 - 1)2 +(0 – (1/2)2 = **√** 9+1/4 =

D(A,e3) = **√** (-2 - (-2,5))2 +(0 – 0)2 = **√** 0,25 = 0,5

Минимальным является расстояние d(A,e3)

Сравниваем расстояние от точки B(-2; -1) до эталонных точек.

D(B,e1) = **√** (-2 – 7/3)2 +(-1 – 0)2 = **√** ((13/3)2 + 1) =

D(B,e2) = **√** (-2 - (-2))2 +(-1 – (-1))2 = **√** (0 + 0) = **0**

D(B,e3) = **√** (-2 - (-2,5))2 +(-1 – 0)2 =

Минимальным является расстояние d(B,e2)

Сравниваем расстояние от точки C(-3; 0) до эталонных точек.

D(C,e1) = **√** (-3 – 7/3)2 +(0 – 0)2 = **√** (16/3)2

D(C,e2) = **√** (-3 - (-2))2 +(0 – (-1))2 = **√** (1 + 1) = **√** 2

D(C,e3) = **√** (-3 - (-2,5))2 +(0 – 0)2 = **√** 0,25+0 = **√** 0,25

Минимальным является расстояние d(C,e3)

Сравниваем расстояние от точки D(4; 0) до эталонных точек.

D(D,e1) = **√** (4 – 7/3)2 +(0 – 0)2 = **√** (5/3)2 =

D(D,e2) = **√** (4 - (-2))2 +(0 – (-1))2 = **√** (36 + 1) = **√** 37

D(D,e3) = **√** (4 - (-2,5))2 +(0 – 0)2 = **√** 39,0625

Минимальным является расстояние d(D,e1)

Сравниваем расстояние от точки E(3; 0) до эталонных точек.

D(E,e1) = **√** (3 – 7/3)2 +(0 – 0)2 = **√** (2/3)2

D(E,e2) = **√** (3 - (-2))2 +(0 – (-1))2 = **√** (25 + 1) = **√** 26

D(E,e3) = **√** (3 - (-2,5))2 +(0 – 0)2 = **√** 30,25

Минимальным является расстояние d(E,e1)

Сравниваем расстояние от точки F(0; 0) до эталонных точек.

D(F,e1) = **√** (0 – 7/3)2 +(0 – 0)2 = 7/3

D(F,e2) = **√** (0 - (-2))2 +(0 – (-1))2 = **√** (1 + 0,25) = **√** 1,25

D(F,e3) = **√** (0 - (-2,5))2 +(0 – 0)2 = **√** 9+0 = 3

Минимальным является расстояние d(F,e2)

Выписываем в таблице принадлежность объектов кластерам

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Объект | A | B | C | D | E | F |
| Номер кластера | 3 | 2 | 3 | 1 | 1 | 2 |

|  |  |  |
| --- | --- | --- |
| e1 | e2 | e3 |
| D, E | B, F | A, C |

Пересчитываем значения для эталонных точек.

e1: (4+3)/2 = 3,5; (0+0)/2 = 0; e1 =(3,5;0)

e2: (-2 +0)/2 = -1; (-1+0)/2 = -0,5; e2 =(-1; -0,5)

e3: (-2 -3)/2 = -2,5; (0+0)/2 = 0; e3 =(-2,5;0)

Границы кластеров изменились, продолжаем процесс разбиения.

**Итерация №4**

Эталонные точки - центроиды кластеров:

e1 = (3,5; 0)

e2 = (-1; -0,5)

e3 = (-2,5; 0)

Сравниваем расстояние от точки A(-2;0) до эталонных точек.

D(A,e1) = **√** (-2 – 3,5)2 +(0 – 0)2 = **√** (-5,5)2 = **√** 27,5

D(A,e2) = **√** (-2 - (-1))2 +(0 – (-0,5)2 = **√** (1+0,25)= **√** 1,25

D(A,e3) = **√** (-2 - (-2,5))2 +(0 – 0)2 = **√** 0,25 = 0,5

Минимальным является расстояние d(A,e3)

Сравниваем расстояние от точки B(-2; -1) до эталонных точек.

D(B,e1) = **√** (-2 – 3,5)2 +(-1 – 0)2 = **√** ((-5,5)2 + 1) = **√** 31,25

D(B,e2) = **√** (-2 - (-1))2 +(-1 – (-0,5))2 = **√** (1 + 0,25) = **√** 1,25

D(B,e3) = **√** (-2 - (-2,5))2 +(-1 – 0)2 = **√** 0,25+1= **√** 1,25

Минимальным является расстояние d(B,e2)

Сравниваем расстояние от точки C(-3; 0) до эталонных точек.

D(C,e1) = **√** (-3 – 3,5)2 +(0 – 0)2 = **√** (-7,5)2 =7,5

D(C,e2) = **√** (-3 - (-1))2 +(0 – (-0,5))2 = **√** (4 + 0,25) = **√** 4,25

D(C,e3) = **√** (-3 - (-2,5))2 +(0 – 0)2 = **√** 0,25+0 = **√** 0,25

Минимальным является расстояние d(C,e3)

Сравниваем расстояние от точки D(4; 0) до эталонных точек.

D(D,e1) = **√** (4 – 3,5)2 +(0 – 0)2 = **√** (0,5)2 = **√** 0,25 = 0,5

D(D,e2) = **√** (4 - (-1))2 +(0 – (-0,5))2 = **√** (25 + 0,25) = **√** 25,25

D(D,e3) = **√** (4 - (-2,5))2 +(0 – 0)2 = **√** 39,0625

Минимальным является расстояние d(D,e1)

Сравниваем расстояние от точки E(3; 0) до эталонных точек.

D(E,e1) = **√** (3 – 3,5)2 +(0 – 0)2 = **√** 0,25= 0,5

D(E,e2) = **√** (3 - (-1))2 +(0 – (-0,5))2 = **√** (16 + 0,25) = **√** 16,25

D(E,e3) = **√** (3 - (-2,5))2 +(0 – 0)2 = **√** 30,25

Минимальным является расстояние d(E,e1)

Сравниваем расстояние от точки F(0; 0) до эталонных точек.

D(F,e1) = **√** (0 – 3,5)2 +(0 – 0)2 = **√** ( 3,5)2

D(F,e2) = **√** (0 - (-1))2 +(0 – (-0,5))2 = **√** (1 + 0,25) = **√** 1,25

D(F,e3) = **√** (0 - (-2,5))2 +(0 – 0)2 = **√** 9+0 = 3

Минимальным является расстояние d(F,e2)

Выписываем в таблице принадлежность объектов кластерам

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Объект | A | B | C | D | E | F |
| Номер кластера | 3 | 2 | 3 | 1 | 1 | 2 |

|  |  |  |
| --- | --- | --- |
| e1 | e2 | e3 |
| D, E | B, F | A, C |

Пересчитываем значения для эталонных точек.

e1: (4+3)/2 = 3,5; (0+0)/2 = 0; e1 =(3,5;0)

e2: (-2 +0)/2 = -1; (-1+0)/2 = -0,5; e2 =(-1; -0,5)

e3: (-2 -3)/2 = -2,5; (0+0)/2 = 0; e3 =(-2,5;0)

Границы кластеров не изменились, процесс разбиения на кластеры закончен.