Методы многомерной классификации

- Тема 4. **Кластерный анализ**
- Тема 5. **Дискриминантный анализ.**

Основные понятия

- **Классификация** – разделение рассматриваемой совокупности объектов или явлений на однородные, в определенном смысле, группы либо отнесение каждого объекта (из заданного множества) к одному из заранее известных классов.

Исходная информация

1. Пусть каждый *i*-ый из анализируемых объектов $(i=\overline{1,n})$ характеризуется значениями определенного набора признаков (свойств) $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(p)}$, т.е. речь идет о классификации многомерных наблюдений $x_1, x_2, \ldots, x_n$, где $x_i = (x_i^{(1)}, \ldots, x_i^{(p)})^T$ - в-р столбец значений признаков для i-го объекта.
2. Обучающие выборки $X_{j1}, X_{j2}, \ldots, X_{jn}$, j = 1, 2,…, k, каждая (*j*-ая) из которых определяет значения анализируемых признаков $X_{ji} = (x_{ji}^{(1)}, x_{ji}^{(2)}, \ldots, x_{ji}^{(p)})^T$ на n объектах, о которых априори известно, что они принадлежат j-му классу, причем число k различных выборок равно общему числу всех

Результат

- 1. если число классов k и их смысл **известны заранее**, то каждое из n классифицируемых многомерных наблюдений должно быть снабжено «адресом» (номером) класса, к которому оно принадлежит.

или

- 2. если число классов k и (или) их смысл **выявляются в процессе классификации**, то результатом классификации является разделение множества классифицируемых объектов на определенное число однородных (в определенном смысле) групп, каждая из которых объявляется классом.

- Если в начале располагают не только классифицируемыми данными, но и обучающими выборками, то имеем задачу классификации при наличии обучающих выборок или **«классификации с обучением» (методы дискриминантного анализа)**;
- в противном случае – задача **«классификации без обучения» (методы кластерного анализа)**.

Кластерный анализ

- **Постановка задачи.**

Пусть исследуется совокупность n объектов, каждый из которых характеризуется по k замеренным на нем признакам X.

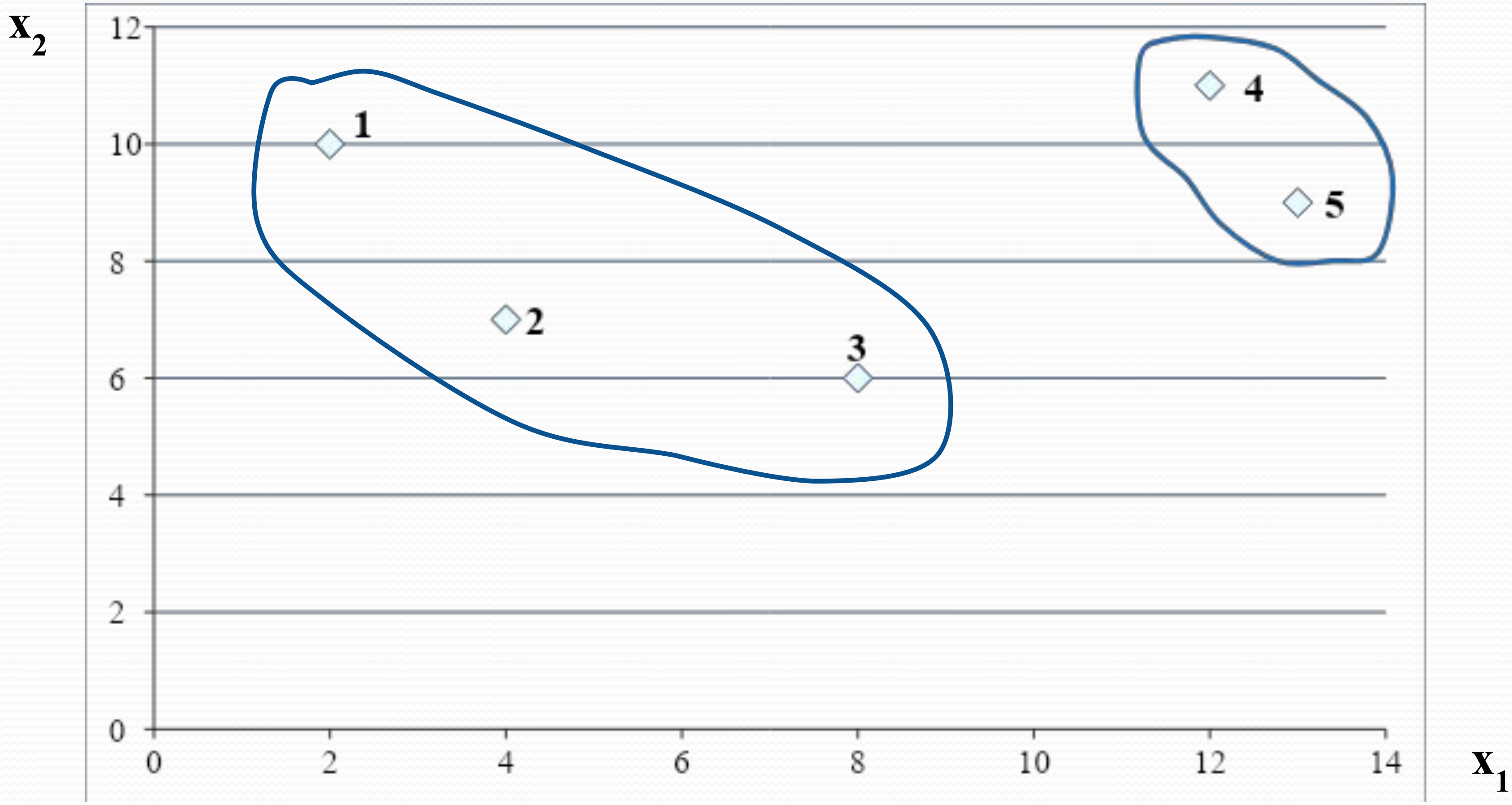
Требуется разбить эту совокупность на однородные в некотором смысле группы (классы).

Полученные в результате разбиения группы обычно называются **кластерами**.

Расстояние между *объектами* и мера близости

- **Понятие однородности объектов**

задается либо введением правила вычисления расстояния $\rho(X_i, X_j)$ между любой парой исследуемых объектов $(X_1, X_2, \ldots, X_n)$, либо заданием некоторой функции $r(X_i, X_j)$, характеризующей степень близости i-го и j-го объектов.



x2
12
10
8
6
4
2
0
1
2
3
4
5
0
2
4
6
8
10
12
14
x1

Расстояние между <u>*объектами*</u> и мера близости

- **обычное евклидово расстояние;**

$$\rho_E(X_i, X_j) = \sqrt{\sum_{l=1}^{k} (x_{il} - x_{jl})^2}$$

x_{il} – значение *l*-го признака у *i*-го объекта.

Для приведения признаков к одинаковым единицам прибегают к нормировке каждого признака:

$$x_{il}^{H} = \frac{x_{il} - \bar{x}_l}{\sigma_l}$$

Расстояние между ***объектами*** и мера близости (продолжение)

- **«взвешенное» евклидово расстояние;**

$$\rho_{BE}(X_i, X_j) = \sqrt{\sum_{l=1}^{k} w_l (x_{il} - x_{jl})^2}$$

Каждой компоненте x_l вектора X приписывают некоторый вес w_l, учитывающий степень важности данного признака в задаче классификации.

Обычно полагают $0 \le w_l \le 1$, где $l = 1, 2, \dots, k$.

Расстояние между *объектами* и мера близости (продолжение)

- **хеммингово расстояние;**

$$\rho_H(X_i, X_j) = \sum_{l=1}^{k} | x_{il} - x_{jl} |$$

Используется как мера различия объектов, задаваемых дихотомическими признаками. Хеммингово расстояние равно числу несовпадений значений соответствующих признаков в рассматриваемых *i*-ом и *j*-ом объектах.

Расстояние между *кластерами*

Пусть S_i – i-я группа (кластер), состоящая из n_i объектов.

Способы вычисления расстояния между кластерами:

1. расстояние по принципу **«ближайшего соседа»**:

$$\rho_{\min}(S_l, S_m) = \min_{\substack{x_i \in S_l \\ x_j \in S_m}} \rho\left(x_i, x_j\right)$$

Расстояние между кластерами

2. расстояние по принципу **«дальнего соседа»**:

$$\rho_{\max}(S_l, S_m) = \max_{\substack{x_i \in S_l \\ x_j \in S_m}} \rho\left(x_i, x_j\right)$$

3. расстояние между **центрами тяжести классов**

$$\rho(S_l, S_m) = \rho\left(\overline{x}_l, \overline{x}_m\right)$$

где $\overline{x}_l = \dfrac{1}{n_l} \sum_{x_i \in S_l} x_i$ - «центр тяжести» l-ой группы.

Расстояние между кластерами

4. расстояние по принципу **«средней связи»**:

$$\rho_{cp}(S_l, S_m) = \frac{1}{n_l n_m} \cdot \sum_{x_i \in S_l} \sum_{x_j \in S_m} \rho\left(x_i, x_j\right)$$

- среднее арифметическое всех попарных расстояний между представителями рассматриваемых групп.

Иерархические процедуры классификации

Иерархические процедуры классификации позволяют получить представление о структуре классифицируемой совокупности наблюдений в форме дендрограммы.

Различают **агломеративные** и **дивизимные** иерархические процедуры классификации .

Иерархические процедуры классификации

Агломеративные иерархические процедуры классификации на первом шаге рассматривают каждое из классифицируемых наблюдений $x_i \, (i = \overline{1, n})$ как отдельный кластер.

Далее на каждом шаге алгоритма происходит объединение двух самых близких наблюдений, затем – двух самых близких групп наблюдений (кластеров). Работа алгоритма заканчивается, когда все исходящие наблюдения оказались объединенными в один класс.

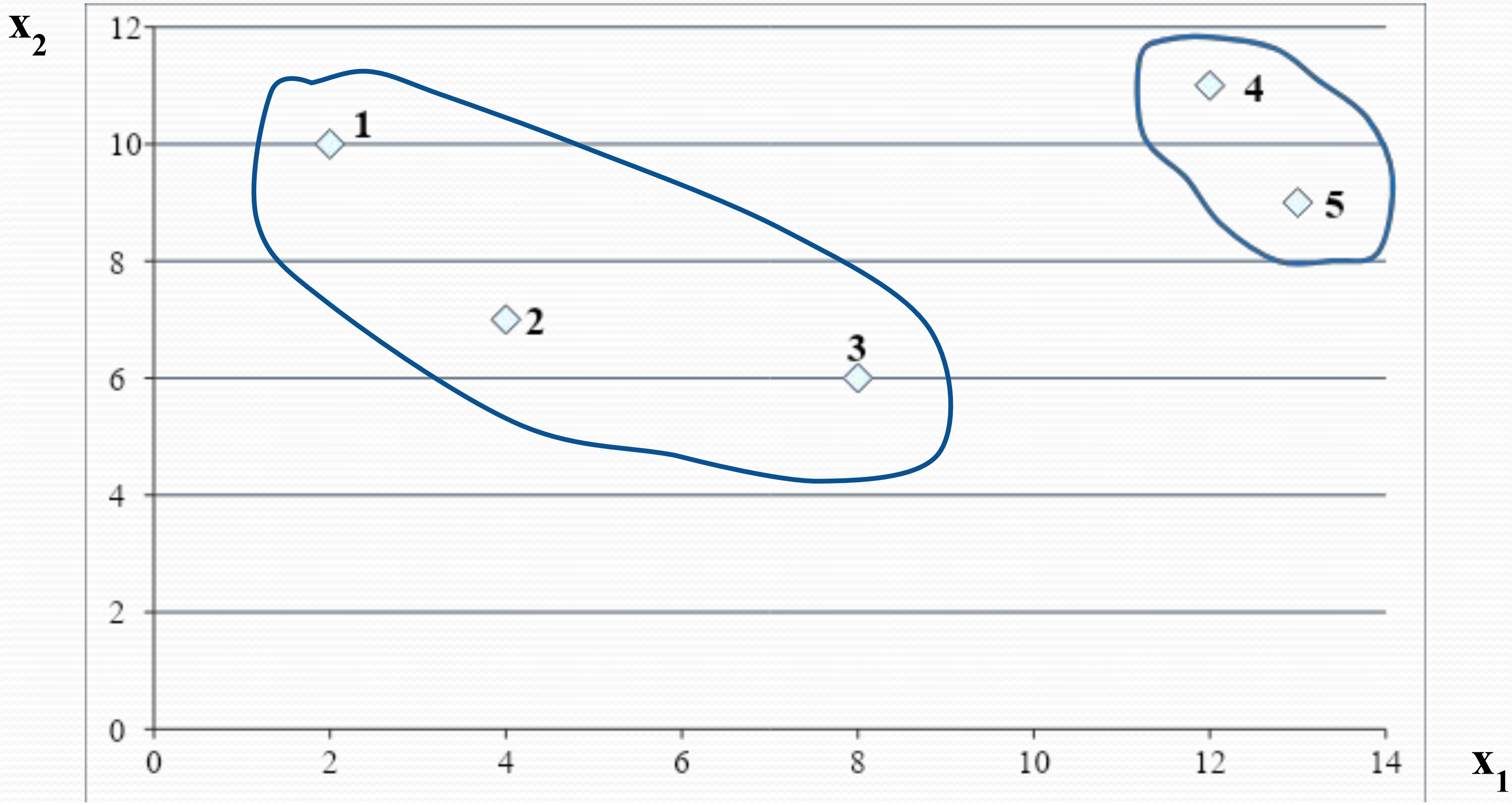
Иерархические процедуры классификации

ПРИМЕР

По приведенным данным провести классификацию 5 семей по двум показателям: уровень расходов за летние месяцы на культурные нужды, спорт и отдых (x_1), уровень расходов на питание (x_2).

№ семьи	1	2	3	4	5
x_{i1}	2	4	8	12	13
x_{i2}	10	7	6	11	9

Иерархические процедуры классификации



Иерархические процедуры классификации

РЕШЕНИЕ

1.1. Евклидова метрика. Принцип «ближнего соседа».

$$\rho_{1,2} = \sqrt{\sum_{l=1}^{k} (x_{1l} - x_{2l})^2} = \sqrt{(2-4)^2 + (10-7)^2} = 3{,}61$$

$$\rho_{1,3} = \sqrt{\sum_{l=1}^{k} (x_{1l} - x_{3l})^2} = \sqrt{(2-8)^2 + (10-6)^2} = 7{,}21$$

$$\rho_{1,1} = 0$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$R_1 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 & 5{,}83 \\ 10{,}05 & 8{,}94 & 6{,}40 & 0 & 2{,}24 \\ 11{,}05 & 9{,}22 & 5{,}83 & 2{,}24 & 0 \end{pmatrix} \quad \rho_{4,5}$$

Имеем: S_1, S_2, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами S_1 и $S_{4,5}$ равно:

$$\rho_{\min}(S_1, S_{4,5}) = \frac{1}{2}\rho_{1,4} + \frac{1}{2}\rho_{1,5} - \frac{1}{2}\left|\rho_{1,4} - \rho_{1,5}\right| =$$

$$= \frac{1}{2}(10,05 + 11,05) - \frac{1}{2}\left|10,05 - 11,05\right| = 10,05$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$\rho_{1,2} \qquad \rho_{1,(4,5)} \qquad \rho_{2,(4,5)}$$

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 \\ 7{,}21 & 4{,}12 & 0 & 5{,}83 \\ 10{,}05 & 8{,}94 & 5{,}83 & 0 \end{pmatrix}$$

Имеем: $S_{1,2}$, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами $S_{1,2}$ и $S_{4,5}$ равно:

$$\rho_{\min}(S_{4,5}, S_{1,2}) = \frac{1}{2}\rho_{(4,5),1} + \frac{1}{2}\rho_{(4,5),2} - \frac{1}{2}\left|\rho_{(4,5),1} - \rho_{(4,5),2}\right| =$$

$$= \frac{1}{2}(10{,}05 + 8{,}94) - \frac{1}{2}\left|10{,}05 - 8{,}94\right| = 8{,}94$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$\rho_{(1,2),3}$

$$R_3 = \begin{pmatrix} 0 & 4{,}12 & 8{,}94 \\ 4{,}12 & 0 & 5{,}83 \\ 8{,}94 & 5{,}83 & 0 \end{pmatrix}$$

Имеем: $S_{1,2,3}$, $S_{4,5}$

$$R_4 = \begin{pmatrix} 0 & 5{,}83 \\ 5{,}83 & 0 \end{pmatrix}$$

Иерархические процедуры классификации

Иерархические процедуры классификации

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

1.2. Принцип «дальнего соседа».

$$R_1 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 & 5{,}83 \\ 10{,}05 & 8{,}94 & 6{,}40 & 0 & 2{,}24 \\ 11{,}05 & 9{,}22 & 5{,}83 & 2{,}24 & 0 \end{pmatrix} \longleftarrow \rho_{4,5}$$

Имеем: $S_1, S_2, S_3, S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами S_1 и $S_{4,5}$ равно:

$$\rho_{\max}(S_1, S_{4,5}) = \frac{1}{2}\rho_{1,4} + \frac{1}{2}\rho_{1,5} + \frac{1}{2}\left|\rho_{1,4} - \rho_{1,5}\right| =$$

$$= \frac{1}{2}(10{,}05 + 11{,}05) + \frac{1}{2}\left|10{,}05 - 11{,}05\right| = 11{,}05$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$\rho_{1,2}$ $\rho_{1,(4,5)}$ $\rho_{2,(4,5)}$

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 \\ 11{,}05 & 9{,}22 & 6{,}40 & 0 \end{pmatrix}$$

Имеем: $S_{1,2}$, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами $S_{1,2}$ и $S_{4,5}$ равно:

$$\rho_{\max}(S_{4,5}, S_{1,2}) = \frac{1}{2}\rho_{(4,5),1} + \frac{1}{2}\rho_{(4,5),2} + \frac{1}{2}\left|\rho_{(4,5),1} - \rho_{(4,5),2}\right| =$$

$$= \frac{1}{2}(11{,}05 + 9{,}22) + \frac{1}{2}\left|11{,}05 - 9{,}22\right| = 11{,}05$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$R_3 = \begin{pmatrix} 0 & 7{,}21 & 11{,}05 \\ 7{,}21 & 0 & 6{,}40 \\ 11{,}05 & 6{,}40 & 0 \end{pmatrix}$$

$\rho_{3,(4,5)}$

Имеем: $S_{1,2}$, $S_{3,4,5}$

$$R_4 = \begin{pmatrix} 0 & 11{,}05 \\ 11{,}05 & 0 \end{pmatrix}$$

Иерархические процедуры классификации

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

1.3. Принцип «центра тяжести».

$$R_1 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 & 5{,}83 \\ 10{,}05 & 8{,}94 & 6{,}40 & 0 & 2{,}24 \\ 11{,}05 & 9{,}22 & 5{,}83 & 2{,}24 & 0 \end{pmatrix} \leftarrow \rho_{4,5}$$

Имеем: $S_1, S_2, S_3, S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Кластер $S_{4,5}$ характеризуется центром тяжести:

$$\overline{X}_{4,5} = \begin{pmatrix} 12{,}5 \\ 10 \end{pmatrix}$$

Расстояние между кластерами S_1 и $S_{4,5}$ равно:

$$\rho_{(4,5),1} = \sqrt{(12{,}5-2)^2 + (10-10)^2} = 10{,}5$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$\rho_{1,2}$

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 \\ 3{,}61 & 0 & 4{,}12 & 9{,}01 \\ 7{,}21 & 4{,}12 & 0 & 6{,}02 \\ 10{,}05 & 9{,}01 & 6{,}02 & 0 \end{pmatrix}$$

Имеем: $S_{1,2}$, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Кластер $S_{1,2}$ характеризуется центром тяжести:

$$\overline{X}_{1,2} = \begin{pmatrix} 3 \\ 8{,}5 \end{pmatrix}$$

Расстояние между кластерами $S_{1,2}$ и $S_{4,5}$ равно:

$$\rho_{(4,5),(1,2)} = \sqrt{(3-12{,}5)^2 + (8{,}5-10)^2} = 9{,}62$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$\rho_{3,(1,2)}$$

$$R_3 = \begin{pmatrix} 0 & 5{,}59 & 9{,}62 \\ 5{,}59 & 0 & 6{,}02 \\ 9{,}62 & 6{,}02 & 0 \end{pmatrix}$$

Имеем: $S_{1,2,3}$, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Кластер $S_{1,2,3}$ характеризуется центром тяжести:

$$\overline{X}_{1,2,3} = \begin{pmatrix} (2+4+8)/3 \\ (10+7+6)/3 \end{pmatrix} = \begin{pmatrix} 4,67 \\ 7,67 \end{pmatrix}$$

Расстояние между кластерами $S_{1,2,3}$ и $S_{4,5}$ равно:

$$\rho_{(1,2,3),(4,5)} = \sqrt{(4,67-12,5)^2 + (7,67-10)^2} = 8,17$$

$$R_4 = \begin{pmatrix} 0 & 8,17 \\ 8,17 & 0 \end{pmatrix}$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

1.4. Принцип «средней связи».

$$R_1 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}05 & 11{,}05 \\ 3{,}61 & 0 & 4{,}12 & 8{,}94 & 9{,}22 \\ 7{,}21 & 4{,}12 & 0 & 6{,}40 & 5{,}83 \\ 10{,}05 & 8{,}94 & 6{,}40 & 0 & 2{,}24 \\ 11{,}05 & 9{,}22 & 5{,}83 & 2{,}24 & 0 \end{pmatrix} \qquad \rho_{4,5}$$

Имеем: S_1, S_2, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами S_1 и $S_{4,5}$ равно:

$$\rho(S_1, S_{4,5}) = \frac{1}{2}(\rho_{1,4} + \rho_{1,5}) = \frac{1}{2}(10{,}05 + 11{,}05) = 10{,}55$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$\rho_{1,2}$$

$$R_2 = \begin{pmatrix} 0 & 3{,}61 & 7{,}21 & 10{,}55 \\ 3{,}61 & 0 & 8{,}94 & 9{,}08 \\ 7{,}21 & 8{,}94 & 0 & 6{,}12 \\ 10{,}55 & 9{,}08 & 6{,}12 & 0 \end{pmatrix}$$

Имеем: $S_{1,2}$, S_3, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами $S_{1,2}$ и $S_{4,5}$ равно:

$$\rho(S_{1,2}, S_{4,5}) = \frac{1}{4}(\rho_{1,4} + \rho_{1,5} + \rho_{2,4} + \rho_{2,5}) = 9{,}82$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$\rho_{3,(1,2)}$$

$$R_3 = \begin{pmatrix} 0 & 5{,}41 & 9{,}82 \\ 5{,}41 & 0 & 6{,}12 \\ 9{,}82 & 6{,}12 & 0 \end{pmatrix}$$

Имеем: $S_{1,2,3}$, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

Расстояние между кластерами $S_{1,2,3}$ и $S_{4,5}$ равно:

$$\rho(S_{1,2,3}, S_{4,5}) = \frac{1}{6}(\rho_{1,4} + \rho_{1,5} + \rho_{2,4} + \rho_{2,5} + \rho_{3,4} + \rho_{3,5}) = 8{,}58$$

$$R_4 = \begin{pmatrix} 0 & 8{,}58 \\ 8{,}58 & 0 \end{pmatrix}$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

2.1. Взвешенное евклидово расстояние.
Принцип «ближнего соседа».
w_1=0,05 и w_2 = 0,95.

$$\rho_{BE}(X_i, X_j) = \sqrt{\sum_{l=1}^{k} w_l (x_{il} - x_{jl})^2}$$

$$\rho_{1,2} = \sqrt{(2-4)^2 * 0{,}05 + (10-7)^2 * 0{,}95} = 2{,}96$$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$\rho_{2,3}$$

$$R_1 = \begin{pmatrix} 0 & 2{,}96 & 4{,}12 & 2{,}44 & 2{,}65 \\ 2{,}96 & 0 & 1{,}32 & 4{,}29 & 2{,}80 \\ 4{,}12 & 1{,}32 & 0 & 4{,}95 & 3{,}13 \\ 2{,}44 & 4{,}29 & 4{,}95 & 0 & 1{,}96 \\ 2{,}65 & 2{,}80 & 3{,}13 & 1{,}96 & 0 \end{pmatrix}$$

Имеем: S_1, $S_{2,3}$, S_4, S_5

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

$$R_2 = \begin{pmatrix} 0 & 2,96 & 2,44 & 2,65 \\ 2,96 & 0 & 4,29 & 2,80 \\ 2,44 & 4,29 & 0 & 1,96 \\ 2,65 & 2,80 & 1,96 & 0 \end{pmatrix}$$

$\rho_{4,5}$

Имеем: S_1, $S_{2,3}$, $S_{4,5}$

Иерархические процедуры классификации

РЕШЕНИЕ (продолжение)

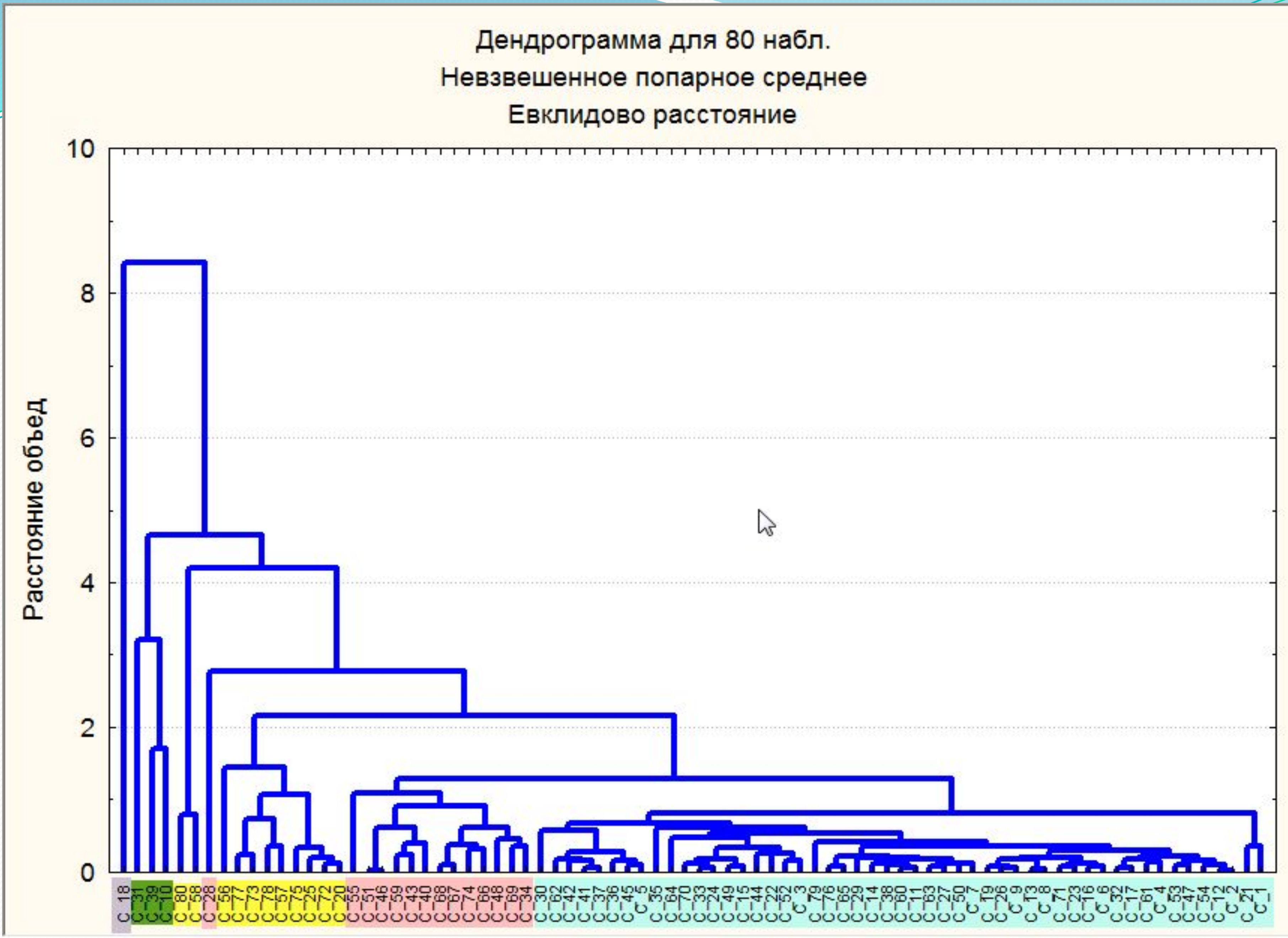
$$\rho_{1,(4,5)}$$

$$R_3 = \begin{pmatrix} 0 & 2{,}96 & 2{,}44 \\ 2{,}96 & 0 & 2{,}80 \\ 2{,}44 & 2{,}80 & 0 \end{pmatrix}$$

Имеем: $S_{1,,4,5}$, $S_{2,3}$

$$R_4 = \begin{pmatrix} 0 & 2{,}80 \\ 2{,}80 & 0 \end{pmatrix}$$

x1 -	**число мест в гостиницах и аналогичных средствах размещения**
x2-	**число мест в специализированных средствах размещения**
x3 -	**объем платных туристских услуг населению, млн. руб.**
x4 -	**объем услуг гостиниц и аналогичных средств размещения, млн. руб.**
x5 -	**объем платных санаторно-оздоровительных услуг населению, млн. руб.**
x6 -	**стоимость турпакетов, реализованных населению, тыс. руб.**
x7 -	**стоимость турпакетов, реализованных гражданам России по территории России, тыс. руб.**
x8 -	**стоимость турпакетов, реализованных гражданам России по зарубежным странам, тыс. руб.**
x9 -	**выручка от оказания туристских услуг, тыс. руб.**
x10 -	**число турфирм, ед.**
x11 -	**число зарегистрированных преступлений, ед.**
x12 -	**среднедушевые денежные доходы населения в месяц, руб.**



Дендрограмма для 80 набл.
Невзвешенное попарное среднее
Евклидово расстояние
Расстояние объед
0
2
4
6
8
10
C_18 C_31 C_39 C_10 C_80 C_58 C_28 C_56 C_77 C_73 C_78 C_57 C_75 C_25 C_72 C_20 C_55 C_51 C_46 C_59 C_43 C_40 C_68 C_67 C_74 C_66 C_48 C_69 C_34 C_30 C_62 C_42 C_41 C_37 C_36 C_45 C_5 C_35 C_64 C_70 C_33 C_24 C_49 C_15 C_44 C_22 C_52 C_3 C_79 C_76 C_65 C_29 C_14 C_38 C_60 C_11 C_63 C_27 C_50 C_7 C_19 C_26 C_9 C_13 C_8 C_71 C_23 C_16 C_6 C_32 C_17 C_61 C_4 C_53 C_47 C_54 C_12 C_2 C_21 C_1

Метод K-средних

Метод *K*-средних предназначен для разбиения многомерных наблюдений на заданное число *K* кластеров, однородных в смысле геометрической взаимной близости элементов, принадлежащих к одному классу.

Метод К-средних

Метод К-средних обычно используется при достаточно больших объемах n классифицируемых данных и реализуется по следующей схеме.

На 1-м этапе производится последовательное (итерационное) уточнение местоположения центров тяжести классов $\overline{x}^{(v)}$ (v – номер итерации).

При этом нулевое приближение $\overline{x}^{(0)}$ строится с помощью случайно выбранных первых K точек исследуемой совокупности, т.е. $\overline{x}^{(0)}_{(i)} = x_i$

Метод K-средних

Затем на j-й итерации «извлекается» точка x_{k+j} и выясняется, к какому из эталонов $\bar{x}^{(0)}_{(i)}$ она оказалась ближе всего. Именно этот, самый близкий к x_{k+j} эталон заменяется новым, определяемым как центр тяжести старого эталона и присоединенной к нему точки x_{k+j}.

Метод K-средних

Имеем нулевое приближение:

$$\bar{x}^{(0)}_{(i)} = x_i$$

$$w^{(0)}_{(i)} = 1\,, \quad i = 1, 2, \ldots, k$$

После «извлечения» новой точки $x_{k+\nu}$ пересчет центров тяжести и весов новых эталонов осуществляется по следующим формулам:

Метод K-средних

$$\bar{x}_i^{(\nu)} = \begin{cases} \dfrac{w_i^{\nu-1}\ \bar{x}_i^{\nu-1} + x_{k+\nu}}{w_i^{\nu-1}+1}, & \text{если} \quad \rho(x_{k+\nu}, \bar{x}_i^{\nu-1}) = \min\limits_{1\le j\le k} \rho(x_{k+\nu}, \bar{x}_j^{\nu-1}) \\ \bar{x}_i^{\nu-1} & \text{в противном случае} \end{cases}$$

$$w_i^{(\nu)} = \begin{cases} w_i^{\nu-1}+1, & \text{если} \quad \rho(x_{k+\nu}, \bar{x}_i^{\nu-1}) = \min\limits_{1\le j\le k} \rho(x_{k+\nu}, \bar{x}_j^{\nu-1}) \\ w_i^{\nu-1} & \text{в противном случае} \end{cases}$$

Метод K-средних

2-й этап метода K-средних посвящен соответственно процедуре классификации.

Окончательное разбиение S исследуемой совокупности на K классов производится в соответствии с правилом «минимального дистанционного разбиения» относительно центров тяжести:

наблюдение x_i относится к классу j_0, если

$$\rho(x_i, \overline{x}_{(j_0)}) = \min_{1 \le j \le k} \rho(x_i, \overline{x}_{(j)})$$

x1 -	**число мест в гостиницах и аналогичных средствах размещения**
x2-	**число мест в специализированных средствах размещения**
x3 -	**объем платных туристских услуг населению, млн. руб.**
x4 -	**объем услуг гостиниц и аналогичных средств размещения, млн. руб.**
x5 -	**объем платных санаторно-оздоровительных услуг населению, млн. руб.**
x6 -	**стоимость турпакетов, реализованных населению, тыс. руб.**
x7 -	**стоимость турпакетов, реализованных гражданам России по территории России, тыс. руб.**
x8 -	**стоимость турпакетов, реализованных гражданам России по зарубежным странам, тыс. руб.**
x9 -	**выручка от оказания туристских услуг, тыс. руб.**
x10 -	**число турфирм, ед.**
x11 -	**число зарегистрированных преступлений, ед.**
x12 -	**среднедушевые денежные доходы населения в месяц, руб.**

Метод K-средних

Элементы кластера номер 3 (Таблица данных34)
и расстояния до центра кластера.
Кластер содержит 28 набл.

Набл.Но. C_2	Набл.Но. C_3	Набл.Но. C_5	Набл.Но. C_7	Набл.Но. C_11	Набл.Но. C_12	Набл.Но. C_14	Набл.Но. C_27	Набл.Но. C_29	Набл.Но. C_30
0,122439	0,194973	0,146170	0,057578	0,116088	0,129804	0,150227	0,065492	0,137372	0,462971

Набл.Но. C_35	Набл.Но. C_36	Набл.Но. C_37	Набл.Но. C_38	Набл.Но. C_41	Набл.Но. C_42	Набл.Но. C_44	Набл.Но. C_45	Набл.Но. C_47	Набл.Но. C_50
0,323378	0,106003	0,202130	0,177634	0,185410	0,145076	0,192376	0,143581	0,141377	0,038825

Набл.Но. C_52	Набл.Но. C_53	Набл.Но. C_60	Набл.Но. C_62	Набл.Но. C_63	Набл.Но. C_64	Набл.Но. C_65	Набл.Но. C_76
0,274909	0,142801	0,141576	0,241598	0,097139	0,405861	0,170628	0,177061

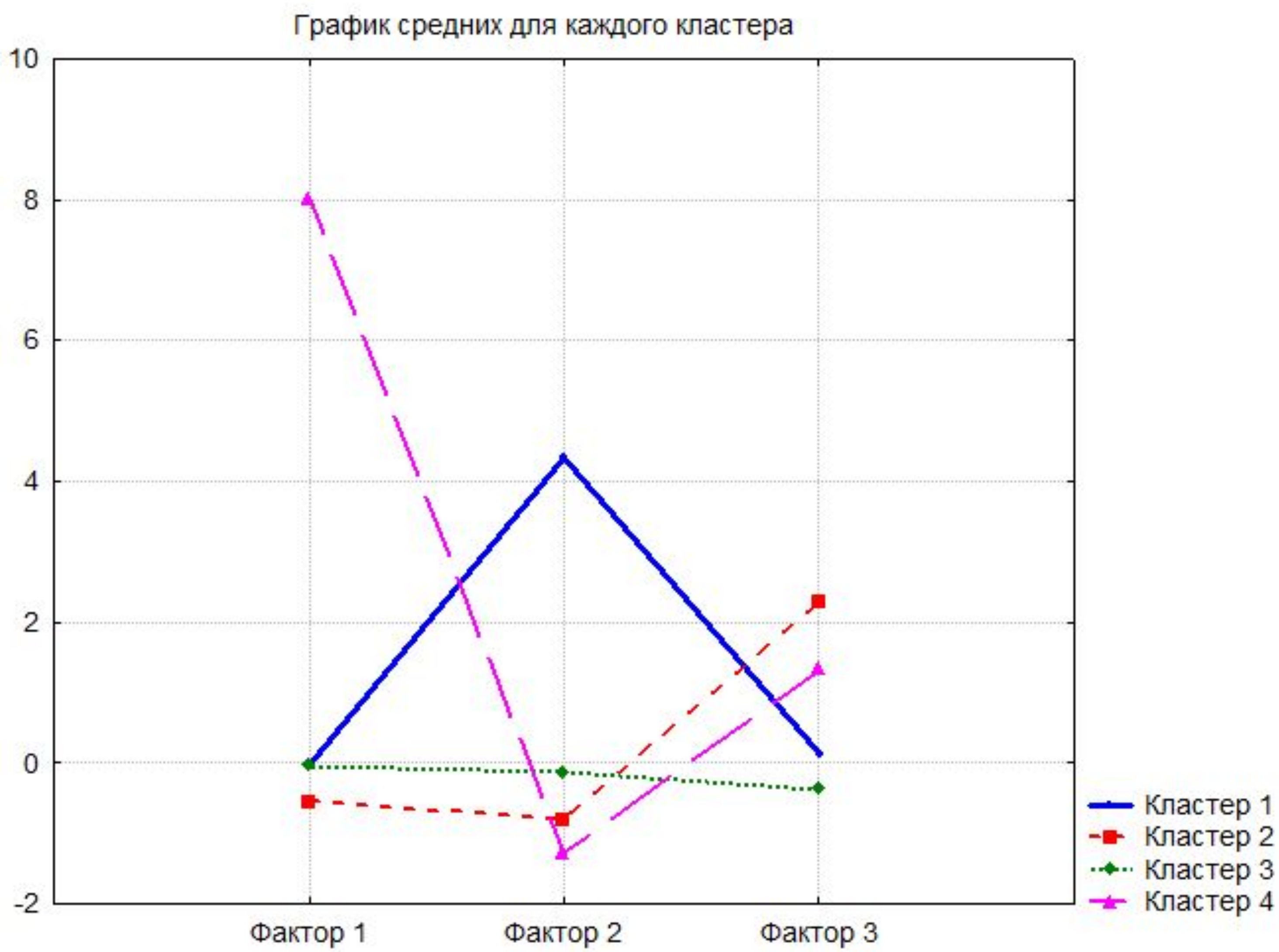


График средних для каждого кластера
10
8
6
4
2
0
-2
Фактор 1
Фактор 2
Фактор 3
Кластер 1
Кластер 2
Кластер 3
Кластер 4