

Практикум

Задача 3.1. Используя данные таблицы 3.1, рассмотрим порядок применения иерархического агломеративного кластерного анализа.

Таблица 3.1 – Данные о потреблении продуктов питания в расчете на одного жителя территории

Виды продуктов питания	Регион A	Регион B	Регион C	Регион D	Регион E
Мясо и мясопродукты на 1 человека в год, кг	59	113	70	94	37
Молоко и молочные продукты на 1 человека в год, кг	363	263	313	412	81
Яйца на 1 человека в год, шт.	268	229	193	257	298
Хлеб и хлебобулочные изделия на 1 человека в год, кг	129	100	90	84	119

Решение: Согласно алгоритму иерархического агломеративного кластерного анализа последовательно проведем указанные процедуры:

1. Сформируем исходную матрицу данных размерностью $n \times m$:

$$X = \begin{Bmatrix} 59 & 363 & 268 & 129 \\ 113 & 263 & 229 & 100 \\ 70 & 313 & 193 & 90 \\ 94 & 412 & 257 & 84 \\ 37 & 81 & 298 & 119 \end{Bmatrix}.$$

2. Проведем нормирование исходных данных по варианту

$$z_{ij} = \frac{x_{ij}}{x_j}:$$

$$x_1 = \frac{59 + 113 + 70 + 94 + 37}{5} = 74,6;$$

$$\begin{aligned}\bar{x}_2 &= 286,4; \\ \bar{x}_3 &= 249,0; \\ \bar{x}_4 &= 104,4.\end{aligned}$$

Тогда

$$\begin{aligned}z_{11} &= \frac{59}{74,6} = 0,79; \\ z_{12} &= \frac{363}{286,4} = 1,27; \\ z_{21} &= \frac{113}{74,6} = 1,51; \\ z_{22} &= \frac{263}{286,4} = 0,92.\end{aligned}$$

В результате матрица нормированных данных Z имеет вид

$$Z = \begin{Bmatrix} 0,79 & 1,27 & 1,08 & 1,24 \\ 1,51 & 0,92 & 0,92 & 0,96 \\ 0,94 & 1,09 & 0,78 & 0,86 \\ 1,26 & 1,44 & 1,03 & 0,80 \\ 0,50 & 0,28 & 1,20 & 1,14 \end{Bmatrix}.$$

3. Используя метрику l -норма: $d_l = \sum_1^m |x_{ik} - x_{jk}|$, перейдем к

матрице расстояний D_0 :

$$\begin{aligned}d_{12} &= |0,79 - 1,51| + |1,27 - 0,92| + |1,08 - 0,92| + |1,24 - 0,96| = 1,51; \\ d_{13} &= |0,79 - 0,94| + |1,27 - 1,09| + |1,08 - 0,78| + |1,24 - 0,86| = 1,01 \text{ и т. д.}\end{aligned}$$

В результате матрица расстояний D_0 имеет вид

$$D_0 = \begin{Bmatrix} 0 & 1,51 & 1,01 & 1,13 & 1,5 \\ 1,51 & 0 & 0,98 & 1,04 & 2,11 \\ 1,01 & 0,98 & 0 & 0,98 & 1,96 \\ 1,13 & 1,04 & 0,98 & 0 & 2,43 \\ 1,50 & 2,11 & 1,96 & 2,43 & 0 \end{Bmatrix}.$$

В агломеративном кластерном анализе могут применяться три алгоритма кластеризации данных: «ближайшего соседа», «дальнего

соседа» и «средней связи». Выберем алгоритм «дальнего соседа», тогда на первом шаге объединим ближайшие объекты $d_{23} = 0,98$ и $d_{34} = 0,98$ и получим следующие кластеры:

Кластер	1	2	3
Объект	1	2, 3, 4	5

Теперь необходимо определить расстояния до второго кластера от всех других объектов, кроме 2, 3, 4 и сформируем матрицу расстояний D_1 :

$$D_1 = \begin{Bmatrix} 0 & 1,51 & 1,50 \\ 1,51 & 0 & 2,43 \\ 1,5 & 2,43 & 0 \end{Bmatrix},$$

минимальное расстояние $d_{13} = 1,50$, следовательно имеем кластеры:

Кластер	1	2
Объект	1, 5	2, 3, 4

Расстояние между двумя кластера равно $d_{12} = 2,43$.

Задачи для самостоятельного решения

Задача 3.2. По данным таблицы 3.2 проведите иерархический агломеративный кластерный анализ.

Таблица 3.2 – Показатели численности и занятости населения за 2006 г. в разрезе районов Ставропольского края

№ п/п	Район	Плотность населения, чел/км ²	Численность безработных, чел.	Численность детей в ДОУ, тыс. чел.	Численность работающих в сельском хозяйстве, тыс. чел.	Численность работающих в промышленном производстве, тыс. чел.
1	Александровский	24,5	1207	1,3	2,0	1,1
2	Андроповский	14,7	738	0,9	0,5	0,9
3	Апанасенковский	9,8	455	1,4	4,7	0,3
4	Арзгирский	8,1	873	1,0	2,5	0,3
5	Благодарненский	26,1	487	2,1	2,2	1,3
6	Буденновский	17,1	230	1,5	3,3	0,3
7	Георгиевский	47,2	1827	2,4	3,4	1,1
8	Грачевский	20,0	1827	1,0	1,2	0,2
9	Изобильненский	51,7	312	3,4	2,7	5,0
10	Ипатовский	16,5	750	1,9	4,3	1,8
11	Кировский	48,0	1255	2,2	1,6	1,2
12	Кочубеевский	33,1	1464	1,7	4,5	2,2
13	Красногвардейский	18,2	1066	1,3	3,2	0,5
14	Курский	14,0	2826	1,2	1,9	0,4
15	Левокумский	9,1	676	1,3	2,5	0,3
16	Минераловодский	32,1	371	4,2	0,9	4,9
17	Нефтекумский	18,3	937	2,1	1,8	4,8
18	Новоалександровский	32,2	512	2,0	4,2	2,2
19	Новоселицкий	14,9	468	0,8	2,5	–
20	Петровский	29,3	1154	2,4	3,3	2,2
21	Предгорный	51,6	678	2,1	2,7	1,3
22	Советский	34,0	1007	2,2	4,1	2,8
23	Степновский	12,1	1464	0,7	2,0	–
24	Труновский	20,6	1190	1,2	3,7	–
25	Туркменский	10,3	406	1,0	1,8	–
26	Шпаковский	46,9	1094	2,9	4,2	2,9

Задача 3.3. По данным таблицы 3.3 проведите иерархический агломеративный кластерный анализ.

Таблица 3.3 – Уровень обеспеченности населения основными продуктами питания за 2006 г. в разрезе районов Ставропольского края

№ п/п	Район	Коэффициент обеспеченности					
		хлеб-ными продуктами	мяс-ными продуктами	молоч-ными продуктами	карто-фелем	сахаром	яйцами
1	Александровский	1,092	0,654	0,465	1,176	1,330	0,901
2	Андроповский	1,102	0,660	0,469	1,187	1,342	0,909
3	Апанасенковский	1,105	0,662	0,470	1,190	1,346	0,911
4	Арзгирский	1,098	0,658	0,467	1,183	1,338	0,906
5	Благодарненский	1,107	0,663	0,471	1,192	1,349	0,913
6	Буденновский	1,109	0,664	0,472	1,194	1,351	0,914
7	Георгиевский	1,194	0,715	0,508	1,286	1,454	0,985
8	Грачевский	1,108	0,664	0,471	1,193	1,350	0,914
9	Изобильненский	1,109	0,665	0,472	1,195	1,352	0,915
10	Ипатовский	1,103	0,661	0,470	1,188	1,344	0,910
11	Кировский	1,117	0,670	0,476	1,204	1,361	0,922
12	Кочубеевский	1,114	0,667	0,474	1,200	1,357	0,918
13	Красногвардейский	1,110	0,665	0,472	1,195	1,352	0,915
14	Курский	1,124	0,674	0,478	1,211	1,370	0,927
15	Левокумский	1,110	0,665	0,473	1,196	1,353	0,916
16	Минераловодский	1,109	0,664	0,472	1,194	1,350	0,914
17	Нефтекумский	1,114	0,668	0,474	1,200	1,357	0,919
18	Новоалександровский	1,101	0,660	0,468	1,186	1,341	0,908
19	Новоселицкий	1,118	0,670	0,476	1,204	1,361	0,922
20	Петровский	1,106	0,663	0,471	1,191	1,347	0,912
21	Предгорный	1,122	0,672	0,478	1,209	1,367	0,925
22	Советский	1,107	0,663	0,471	1,192	1,349	0,913
23	Степновский	1,111	0,666	0,473	1,197	1,354	0,916
24	Труновский	1,098	0,658	0,467	1,183	1,338	0,907
25	Туркменский	1,105	0,662	0,470	1,190	1,346	0,911
26	Шпаковский	1,146	0,687	0,488	1,235	1,396	0,945

Задача 3.4. По данным таблицы 3.4 проведите иерархический агломеративный кластерный анализ.

Таблица 3.4 – Показатели деятельности сельского хозяйства за 2006 г. в разрезе районов Ставропольского края

№ п/п	Район	Прибыль (убыток) сельхозпредприятий, тыс. руб.	Объем с.-х. продукции, произведенной хозяйствами всех категорий – всего, в фактических ценах, млн руб.	Доля растениеводства в общем объеме произведенной с.-х. продукции, %	Производство картофеля во всех категориях хозяйств, тыс. т
1	Александровский	-12587,0	1424,8	63,92	10,4
2	Андроповский	-9867,0	634,8	37,61	10,0
3	Апанасенковский	42305,6	1908,1	51,62	6,0
4	Арзгирский	69602,2	1333,5	69,68	1,0
5	Благодарненский	51125,9	1699,0	66,57	10,1
6	Буденовский	132622,5	2033,2	71,78	4,8
7	Георгиевский	116920,5	2416,2	64,18	11,9
8	Грачевский	-14762,0	1262,1	60,32	7,4
9	Изобильненский	85743,0	2451,6	65,76	41,3
10	Ипатовский	95553,2	2878,3	62,01	0,1
11	Кировский	14578,0	1782,3	62,16	24,6
12	Кочубеевский	141326,5	2902,2	59,01	31,4
13	Красногвардейский	154726,0	3158,0	76,59	15,7
14	Курский	111524,0	1478,8	67,98	1,8
15	Левокумский	116381,0	1969,7	57,98	6,8
16	Минераловодский	15477,2	1219,5	67,55	20,7
17	Нефтекумский	32157,0	1796,9	66,29	9,9
18	Новоалександровский	196126,4	3787,3	76,21	27,1
19	Новоселицкий	201437,0	1878,7	67,04	8,6
20	Петровский	13471,0	2432,1	69,07	16,2
21	Предгорный	66419,4	2248,8	54,11	61,9
22	Советский	289872,2	2496,9	68,75	14,9
23	Степновский	8838,4	1208,6	65,69	1,9
24	Труновский	119266,8	2441,6	62,38	12,3
25	Туркменский	32157,0	1432,1	66,39	1,9
26	Шпаковский	172547,0	4046,8	32,21	30,4

