# Задание

По данным, представленным в таблице, требуется провести классификацию шести семей по двум показателям:

x1 – удельные расходы семьи на летний отдых, %;

x2 – удельные расходы семьи за летние месяцы на культурные нужды и спорт, %.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № п/п | 1 | 2 | 3 | 4 | 5 | 6 |
| x1 | 5 | 6 | 7 | 11 | 9 | 6 |
| x2 | 6 | 4 | 7 | 5 | 7 | 3 |

Классификацию провести по иерархическому агломеративному алгоритмус использованием евклидова расстояния и метода «дальнего соседа»;

# Решение

Воспользуемся агломеративным иерархическим алгоритмом классификации. В качестве расстояния между объектами примем обычное евклидовое расстояние. Тогда согласно формуле:

*=*

Полученные данные помещаем в таблицу (матрицу расстояний).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0,00 | 2,24 | 2,24 | 6,08 | 4,12 | 3,16 |
| 2 | 2,24 | 0,00 | 3,16 | 5,10 | 4,24 | 1,00 |
| 3 | 2,24 | 3,16 | 0,00 | 4,47 | 2,00 | 4,12 |
| 4 | 6,08 | 5,10 | 4,47 | 0,00 | 2,83 | 5,39 |
| 5 | 4,12 | 4,24 | 2,00 | 2,83 | 0,00 | 5,00 |
| 6 | 3,16 | 1,00 | 4,12 | 5,39 | 5,00 | 0,00 |

Поиск наименьшего расстояния.

Из матрицы расстояний следует, что объекты 2 и 6 наиболее близки P2;6 = 1 и поэтому объединяются в один кластер.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0,00 | 2,24 | 2,24 | 6,08 | 4,12 | 3,16 |
| 2 | 2,24 | 0,00 | 3,16 | 5,10 | 4,24 | 1,00 |
| 3 | 2,24 | 3,16 | 0,00 | 4,47 | 2,00 | 4,12 |
| 4 | 6,08 | 5,10 | 4,47 | 0,00 | 2,83 | 5,39 |
| 5 | 4,12 | 4,24 | 2,00 | 2,83 | 0,00 | 5,00 |
| 6 | 3,16 | 1,00 | 4,12 | 5,39 | 5,00 | 0,00 |

При формировании новой матрицы расстояний, выбираем наибольшее значение из значений объектов №2 и №6.

В результате имеем 5 кластера: S(1), S(2,6), S(3), S(4), S(5)

Из матрицы расстояний следует, что объекты 3 и 5 наиболее близки P3;5 = 2 и поэтому объединяются в один кластер.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2,6 | 3 | 4 | 5 |
| 1 | 0,00 | 3,16 | 2,24 | 6,08 | 4,12 |
| 2,6 | 3,16 | 0,00 | 4,12 | 5,39 | 5,00 |
| 3 | 2,24 | 4,12 | 0,00 | 4,47 | 2,00 |
| 4 | 6,08 | 5,39 | 4,47 | 0,00 | 2,83 |
| 5 | 4,12 | 5,00 | 2,00 | 2,83 | 0,00 |

При формировании новой матрицы расстояний, выбираем наибольшее значение из значений объектов №3 и №5.

В результате имеем 4 кластера: S(1), S(2,6), S(3,5), S(4)

Из матрицы расстояний следует, что объекты 1 и 2,6 наиболее близки P1;2,6 = 3.16 и поэтому объединяются в один кластер.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2,6 | 3,5 | 4 |
| 1 | 0,00 | 3,16 | 4,12 | 6,08 |
| 2,6 | 3,16 | 0,00 | 5,00 | 5,39 |
| 3,5 | 4,12 | 5,00 | 0,00 | 4,47 |
| 4 | 6,08 | 5,39 | 4,47 | 0,00 |

При формировании новой матрицы расстояний, выбираем наибольшее значение из значений объектов №1 и №2,6.

В результате имеем 3 кластера: S(1,2,6), S(3,5), S(4)

Из матрицы расстояний следует, что объекты 3,5 и 4 наиболее близки P3,5;4 = 4.47 и поэтому объединяются в один кластер.

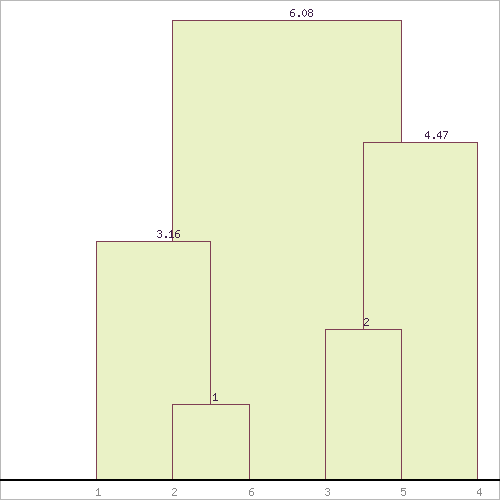
|  |  |  |  |
| --- | --- | --- | --- |
|  | 1,2,6 | 3,5 | 4 |
| 1,2,6 | 0,00 | 5,00 | 6,08 |
| 3,5 | 5,00 | 0,00 | 4,47 |
| 4 | 6,08 | 4,47 | 0,00 |

При формировании новой матрицы расстояний, выбираем наибольшее значение из значений объектов №3,5 и №4.

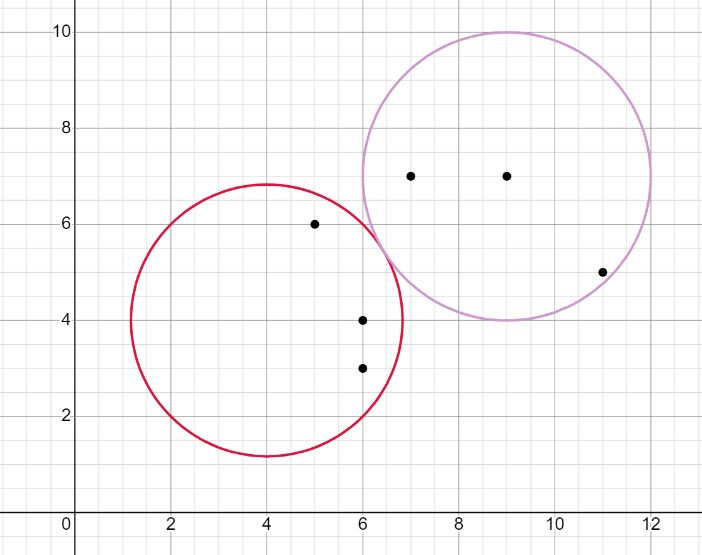
В результате имеем 2 кластера: S(1,2,6), S(3,5,4)

|  |  |  |
| --- | --- | --- |
|  | 1,2,6 | 3,4,5 |
| 1,2,6 | 0,00 | 5,00 |
| 3,4,5 | 5,00 | 0,00 |

Дендрограмма:



Для наглядности результатов классификации представим наблюдения в пространстве переменных x1 и x2:



# Заключение

На основании графического представления результатов кластерного анализа можно сделать вывод, что наилучшим является разбиение шести семей на два кластера: S(1,2,6), S(3,5,4). Таким образом, при проведении кластерного анализа по принципу «Дальний сосед» получили два кластера, расстояние между которыми равно 5,00.

Можно сделать вывод, что семьи 1, 2, 6 и 3, 4, 5 сильно отличаются по летним расходам, поэтому их можно разбить на два кластера соответственно. Первый кластер семей имеет отчетливо меньший расход по обоим параметрам в летнее время: расход на отдых и расход на культурные нужды.