

Алгоритм «Голотип 1»

Назначение

1. Решение задач распознавания в случае, когда в МО представлены объекты только одного образа;
2. Решение задач районирования;
3. Определение представительности МО.

Условия применимости:

ТОС должна быть без пропусков;
свойства — арифметические, логические 1-го и 2-го рода, т.е.
алгоритм может работать с **разношкальными** свойствами.

Постановка задачи

Из предназначения алгоритма вытекают три различные постановки задачи:

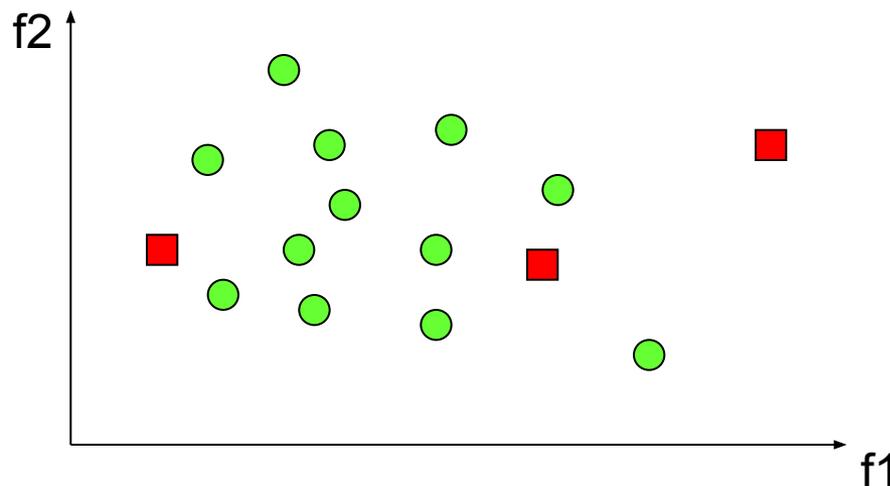
1. Задана совокупность объектов, для части объектов известна принадлежность к образу, относительно оставшейся части объектов следует принять решение о принадлежности их к данному образу (задача распознавания на один образ).

свойства \ объекты	f1	...	f _n	Ψ
a ₁	nn	nn	nn	
...				
a _n	nn	nn	nn	
b ₁	nn	nn	nn	?
...				
b _m	nn	nn	nn	?

- - объекты МО
- - объекты МЭ

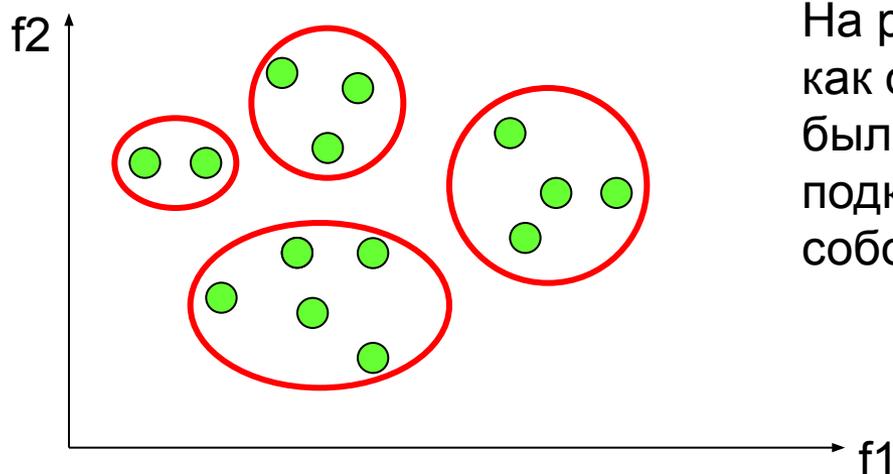
Известно, что объекты a_1, a_2, \dots, a_n относятся к образу (МО), а остальные объекты:

b_1, \dots, b_m - необходимо отнести или не отнести к данному образу.



2. Задана совокупность объектов, которую требуется разбить на группы однородных (в некотором смысле) объектов (задача районирования).

В этом случае изначально нет МЭ, результатом решения такой задачи будут образы (классы) объектов, на которых впоследствии можно решать стандартные задачи распознавания.

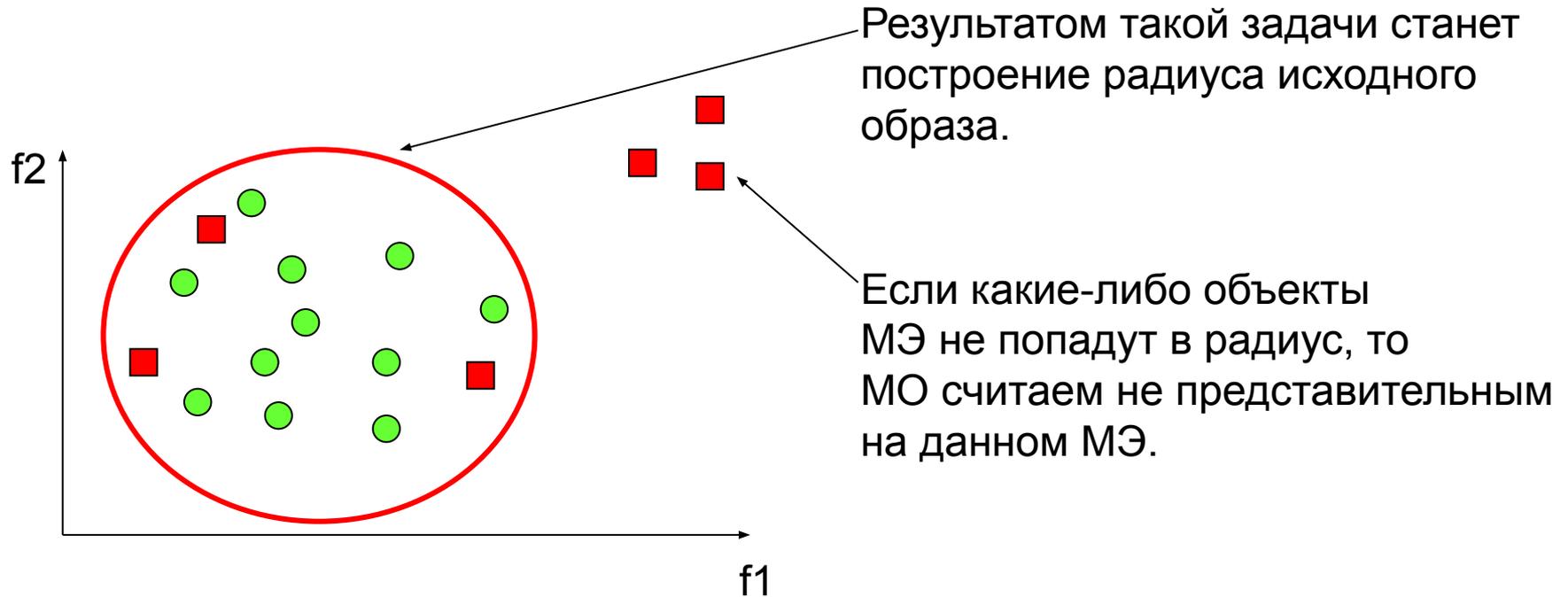


На рисунке образно представлено, как общая совокупность объектов была разбита на некоторое число подклассов (образов) схожих между собой объектов.

● - объекты МО

Алгоритм «Голотип 1»

3. Задана совокупность объектов, для которой необходимо определить представительность МО.



● - объекты МО
■ - объекты МЭ

Решение задачи

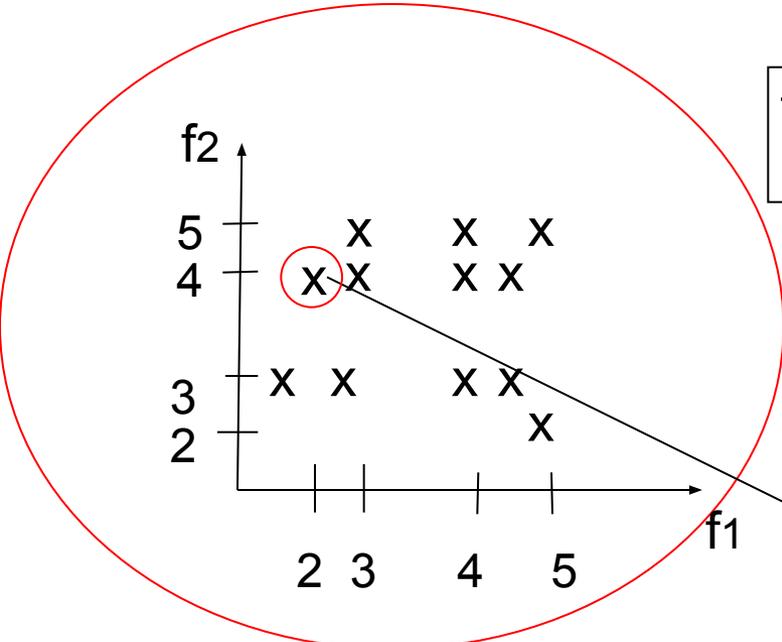
Шаг1: *Постановка задачи*

Для примера рассмотрим задачу районирования:

К примеру нам необходимо распределить совокупность студентов на группы, и из модели мы выбрали два свойства (разумеется в реальной задаче их будут десятки и сотни, но для примера хватит и двух): оценку на экзамене по математике и по английскому. Можем считать, что свойства измерены в логической или арифметической шкале, в данном алгоритме это неважно.

Записываем всех студентов и их оценки в ТОС.

Алгоритм «Голотип 1»



Косвенные свойства

Прямое свойство

ТОС:

	f1	f2	F
a ₁	2	4	1

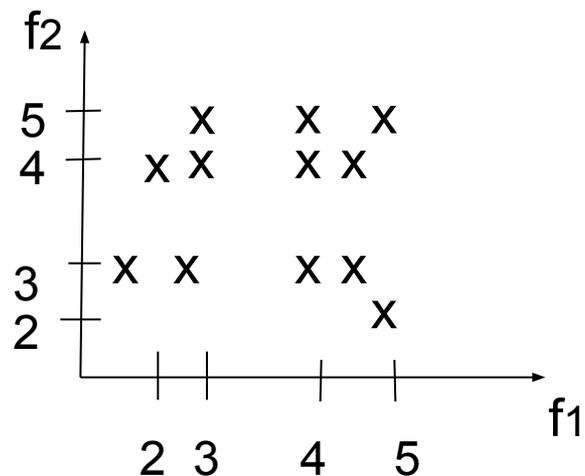
X- студенты
f1- математика
f2- английский

Значения вектора свойств для 1-го объекта

Пространство свойств

Занесем этот объект в таблицу ТОС

Аналогично заносим в ТОС остальные объекты



X- студенты
f1- математика
f2- английский

	f1	f2	F
a1	2	4	I
a2	3	5	I
a3	3	4	I
a4	4	5	I
a5	5	5	I
a6	4	3	I
a7	4	4	I

Шаг2: вычисление экстремальных разностей

По каждому свойству определим минимальное

f_k

$$f_k^* = \min\{f_k\}$$

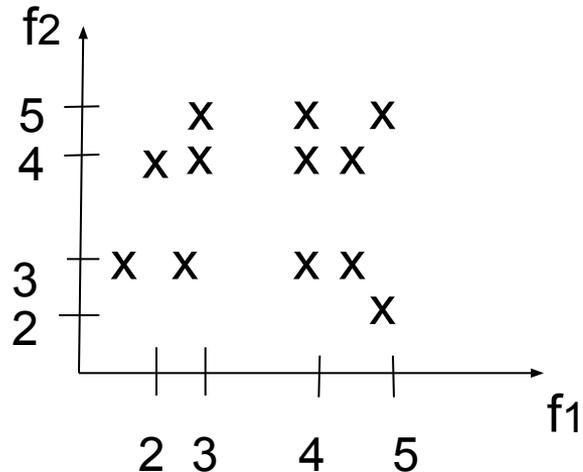
и максимальное

$$f_k^{**} = \max\{f_k\}$$

значения и вычислим экстремальные разности:

$$E_k = f_k^{**} - f_k^*, \text{ где } k = 1, 2, \dots, m$$

m - общее число свойств в ТОС



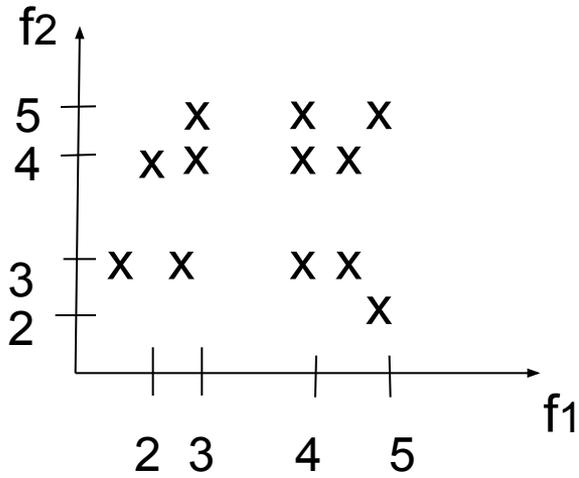
	f1	f2	F
a1	2	4	I
a2	3	5	I
a3	3	4	I
a4	4	5	I
a5	5	5	I
a6	4	3	I
a7	4	4	I

Минимальное значение по свойству f1

Экстремальная разность для 1-го свойства:

$$E_1 = 5 - 2 = 3$$

Максимальное значение по свойству f1



Минимальное значение по свойству f2

Экстремальная разность для 2-го свойства:

$$E_2 = 5 - 3 = 2$$

	f1	f2	F
a1	2	4	I
a2	3	5	I
a3	3	4	I
a4	4	5	I
a5	5	5	I
a6	4	3	I
a7	4	4	I

Максимальное значение по свойству f2

Шаг3: *Вводим меру сходства*

Вычислим меру сходства $\mu_k(i, j)$

между парой объектов a_i, a_j по свойству f_k

1. Если свойство является арифметическим или логическим 2-го рода, то применяется следующая формула:

$$\mu_k(i, j) = 1 - \frac{|f_k^i - f_k^j|}{f_k^{**} - f_k^*}$$

2. Если свойство является логическим 1-го рода, то применяется формула:

$$\mu_k(i, j) = \begin{cases} 1, & \text{если } f_k^i = f_k^j \\ 0, & \text{если } f_k^i \neq f_k^j \end{cases}$$

$$\mu_k(i, j) = 1 - \frac{|f_k^i - f_k^j|}{f_k^{**} - f_k^*}$$

К примеру рассчитаем меру сходства между первым и вторым объектом по первому свойству :

$$\mu_1(1,2) = 1 - \frac{|2-3|}{3-3} = \frac{2}{3} = 0.66$$

Матрица мер сходства

	a1	a2	a3	a4	a5	a6	a7
a1	1						
a2		1					
a3			1				
a4				1			
a5					1		
a6						1	
a7							1

	f1	f2	F
a1	2	4	I
a2	3	5	I
a3	3	4	I
a4	4	5	I
a5	5	5	I
a6	4	3	I
a7	4	4	I

Шаг4:

Рассчитываем матрицы мер сходства для каждого свойства

	f1	f2	F
a1	2	4	1
a2	3	5	1
a3	3	4	1
a4	4	5	1
a5	5	5	1
a6	4	3	1
a7	4	4	1

Матрица мер сходства по 1-му свойству

	a1	a2	a3	a4	a5	a6	a7
a1	1	2/3	2/3	1/3	0	1/3	1/3
a2	2/3	1	1	2/3	1/3	2/3	2/3
a3	2/3	1	1	2/3	1/3	2/3	2/3
a4	1/3	2/3	2/3	1	2/3	1	1
a5	0	1/3	1/3	2/3	1	2/3	2/3
a6	1/3	2/3	2/3	1	2/3	1	1
a7	1/3	2/3	2/3	1	2/3	1	1

Аналогично рассчитывается матрица мер сходства по 2-му свойству

Шаг5: *Вычисляем общую матрицу мер сходства*

по формуле:

$$\mu(i, j) = \sum_{k=1}^m \delta_k \mu_k(i, j) \quad m - \text{число свойств в ТОС};$$

Здесь важно понять смысл коэффициента δ_k (вес k-го свойства)

Он может быть равномерным $\delta_k = \frac{1}{m}$, в случае, если

все свойства равнозначны. Но обычно в реальных задачах именно от подбора этого коэффициента зависит вид решения.

Единственным условием является то, что $\sum_{k=1}^m \delta_k = 1$

$$\mu(i, j) = \sum_{k=1}^2 \frac{1}{2} \mu_k(i, j)$$

$$0,5 * 2/3 + 0,5 * 1 = 5/6 = 0,83$$

Матрица по 1-му свойству

	a1	a2	a3	a4	a5	a6	a7
a1	1	2/3	2/3	1/3	0	1/3	1/3
a2	2/3	1	1	2/3	1/3	2/3	2/3
a3	2/3	1	1	2/3	1/3	2/3	2/3
a4	1/3	2/3	2/3	1	2/3	1	1
a5	0	1/3	1/3	2/3	1	2/3	2/3
a6	1/3	2/3	2/3	1	2/3	1	1
a7	1/3	2/3	2/3	1	2/3	1	1

Общая матрица мер сходства

	a1	a2	a3	a4	a5	a6	a7
a1	1	0,83					
a2		1					
a3			1				
a4				1			
a5					1		
a6						1	
a7							1

	a1	a2	a3	a4	a5	a6	a7
a1	1	1	1/3	1/3	0	2/3	1/3
a2	2/3	1	1	2/3	1/3	2/3	2/3
a3	1/3	1	1	0	1/3	2/3	2/3
a4	1/3	2/3	0	1	2/3	1	1
a5	0	1/3	1/3	2/3	1	2/3	2/3
a6	2/3	2/3	2/3	1	2/3	1	1

Матрица по 2-му свойству
(составлена случайным подбором)

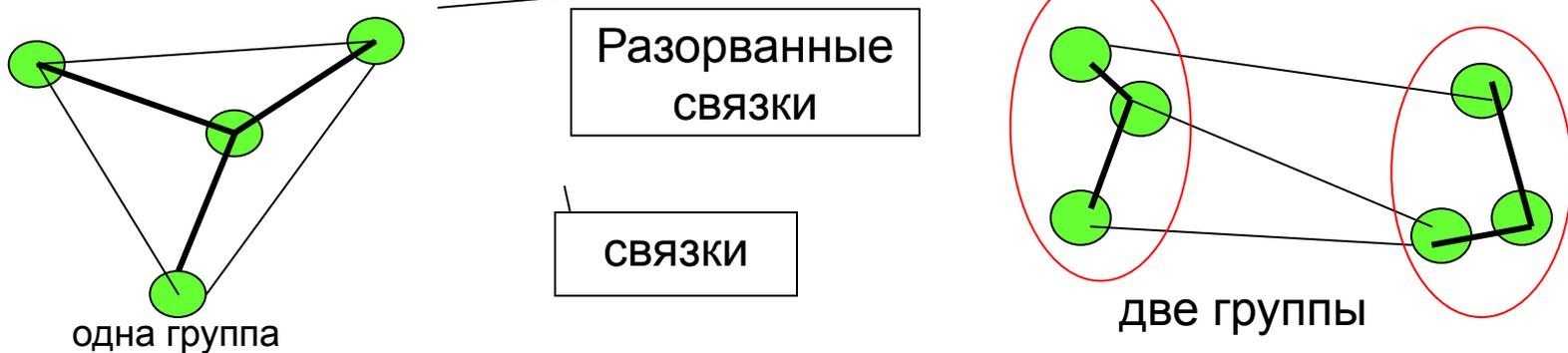
Шаг6: Вычисление порога μ_0

для разбиения материала обучения на однородные группы

Используя меру сходства между объектами по общей матрице мер сходства, можно выделить из исходной совокупности однородные группы, сравнивая меру сходства с пороговым значением μ_0 :

$$\mu(i, j) \geq \mu_0$$

Если условие выполняется, то считаем, что объекты a_i, a_j связаны одной **связкой**. Если нет, то связка рвется, но это не значит, что объекты не могут оказаться в одной группе.



Совокупность объектов $\langle a_i \rangle$, каждый из которых связан с другим посредством любого числа связок, называют **однородной группой**.

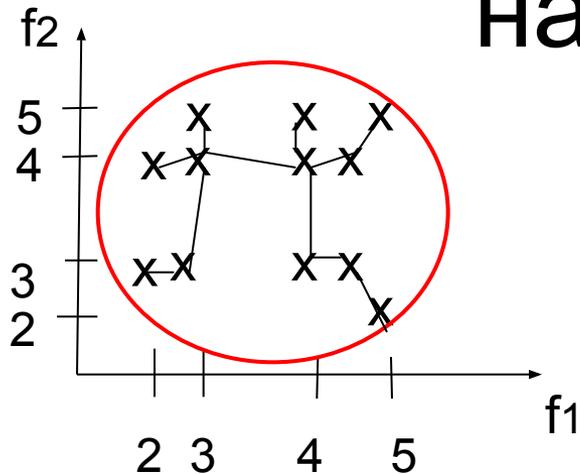
В качестве μ_0 например, выберем среднюю меру сходства по общей матрице мер сходства:

$$\mu_0 = \bar{\mu} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^n \mu(i, j)$$

При подсчете средней меры сходства единицы на диагонали не учитываются!

	a1	a2	a3	a4	a5	a6	a7
a1	1	2/3	2/3	1/3	0	1/3	1/3
a2	2/3	1	1	2/3	1/3	2/3	2/3
a3	2/3	1	1	2/3	1/3	2/3	2/3
a4	1/3	2/3	2/3	1	2/3	1	1
a5	0	1/3	1/3	2/3	1	2/3	2/3
a6	1/3	2/3	2/3	1	2/3	1	1
a7	1/3	2/3	2/3	1	2/3	1	1

Демонстрация влияния μ_0 на вид решения

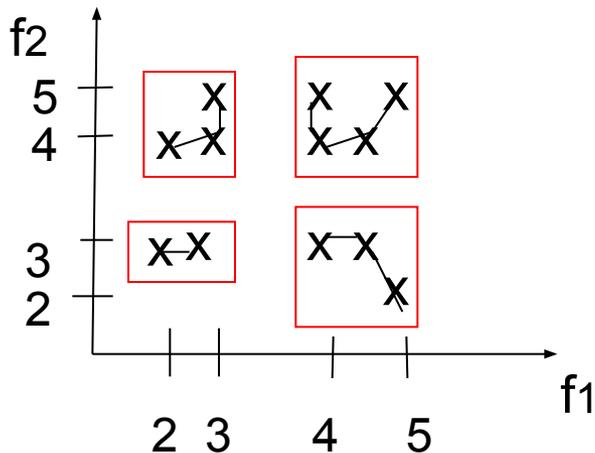


Рассмотрим пример:

1. Если принять $\mu_0 = \bar{\mu}$

т.е. порог равный средней мере сходства между объектами, то может получиться, что все объекты попадут в одну группу.

Это может быть хорошо для задач 1-го типа (задача распознавания на один образ) и 3-го типа (представительность), но плохо для задачи районирования (выделения групп).



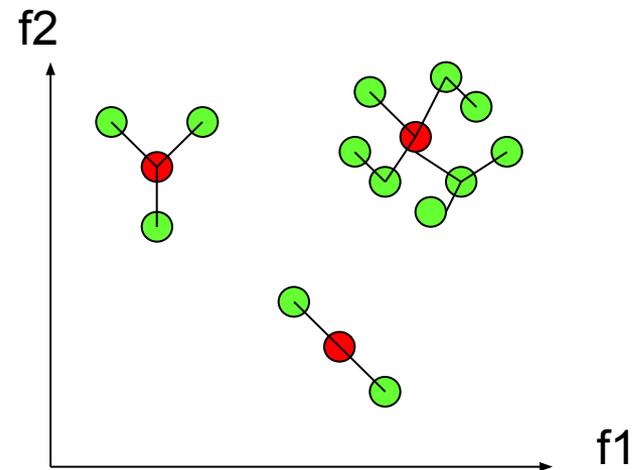
Решается эта проблема путем увеличения порога, чем он больше тем компактнее (однороднее) будут группы, т.к. все «длинные» связки будут разрываться.

Шаг7: Построение голотипов и радиусов (Γ_q, R_q)

Голотип – это тот объект, у которого средняя мера сходства с остальными объектами данной группы является максимальной, т.е. тот на который все остальные объекты в группе наиболее похожи.

$$\Gamma_q = \max_{i=1..m} \left(\frac{\sum_{j=1..m} \mu(i, j)}{m} \right)$$

m – количество объектов в группе

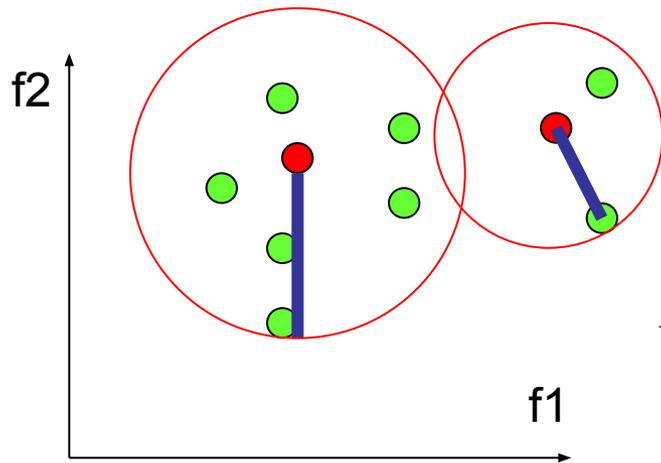


- - объект однородной группы
- - голотип группы

На рисунке представлены примеры голотипов в двумерном пространстве свойств.

Шаг8: Построение радиусов

Радиусом однородной группы является мера сходства голотипа с самым удаленным объектом однородной группы.



Радиус группы равен:

$$R_q = \begin{cases} \mu_0 & , \text{ если в группе один объект;} \\ \min \mu(\Gamma_q, j) & , \text{ если несколько объектов;} \\ 1 \leq j \leq n_q \end{cases}$$

● - объект однородной группы

● - голотип группы

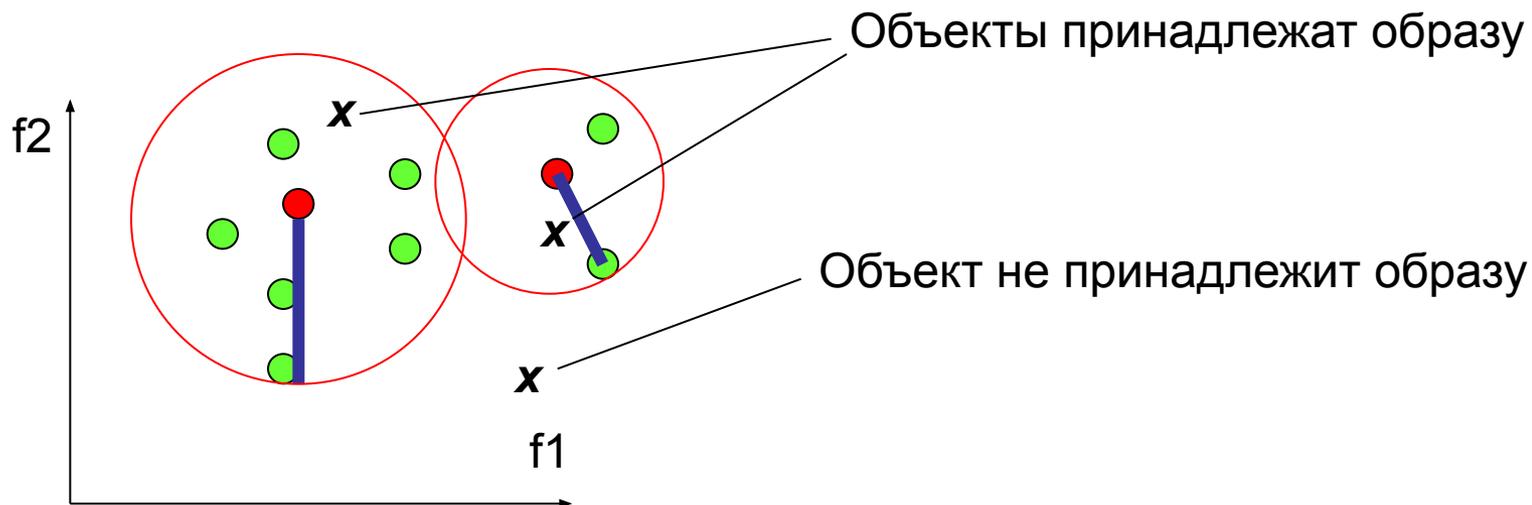
— - радиус группы

Таким образом, однородная группа представляется в виде m - мерного шара, в который входят все объекты, попавшие в данную однородную группу.

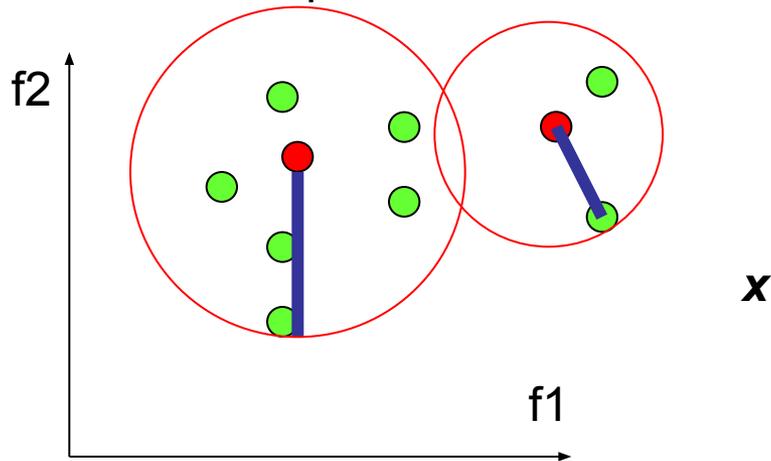
На этом заканчивается процесс обучения и начинается процесс экзамена!

Распознавание объекта « x » производится в соответствии с 3 типами задач следующим образом:

1. Объект принадлежит образу, если существует голотип, мера сходства которого с x больше соответствующего радиуса, в противном случае x к образу не относится.



Для случая исследования материала обучения на представительность для объекта x проводим аналогичные сравнения. МО считается **непредставительным** для x , если нет ни одного голотипа, мера сходства которого с x больше соответствующего ему радиуса.



Для задачи районирования в качестве результата используются все полученные однородные группы.

