

Атрибуты. Кластеризация

Задание №1

Ученый решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звезд — это набор звёзд (точек) на графике, лежащий внутри прямоугольника высотой H и шириной W . Каждая звезда обязательно принадлежит только одному из кластеров.

Истинный центр кластера, или центроид, — это одна из звёзд на графике, сумма евклидовых расстояний от которой до всех остальных звёзд кластера минимальна.

Для двух точек $A(x_1, y_1)$, $B(x_2, y_2)$ евклидово расстояние вычисляется по формуле:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

В файле A хранятся данные о звёздах двух кластеров, где $H=5$, $W=5$ для каждого кластера. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y , затем значение яркости звезды. Значения даны в условных единицах. Известно, что количество звёзд не превышает 1000.

В файле B хранятся данные о звёздах трёх кластеров, где $H=5$, $W=5$ для каждого кластера. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звёздах в файле B аналогична файлу A .

Для каждого файла найдите сумму абсцисс и ординат истинных центроидов кластеров, а затем параметр H — среднее гармоническое значение яркости звезды в каждом кластере.

Среднее гармоническое — это число, которое получается путем взятия обратных величин всех чисел, нахождения среднего арифметического этих величин, а затем обратного преобразования.

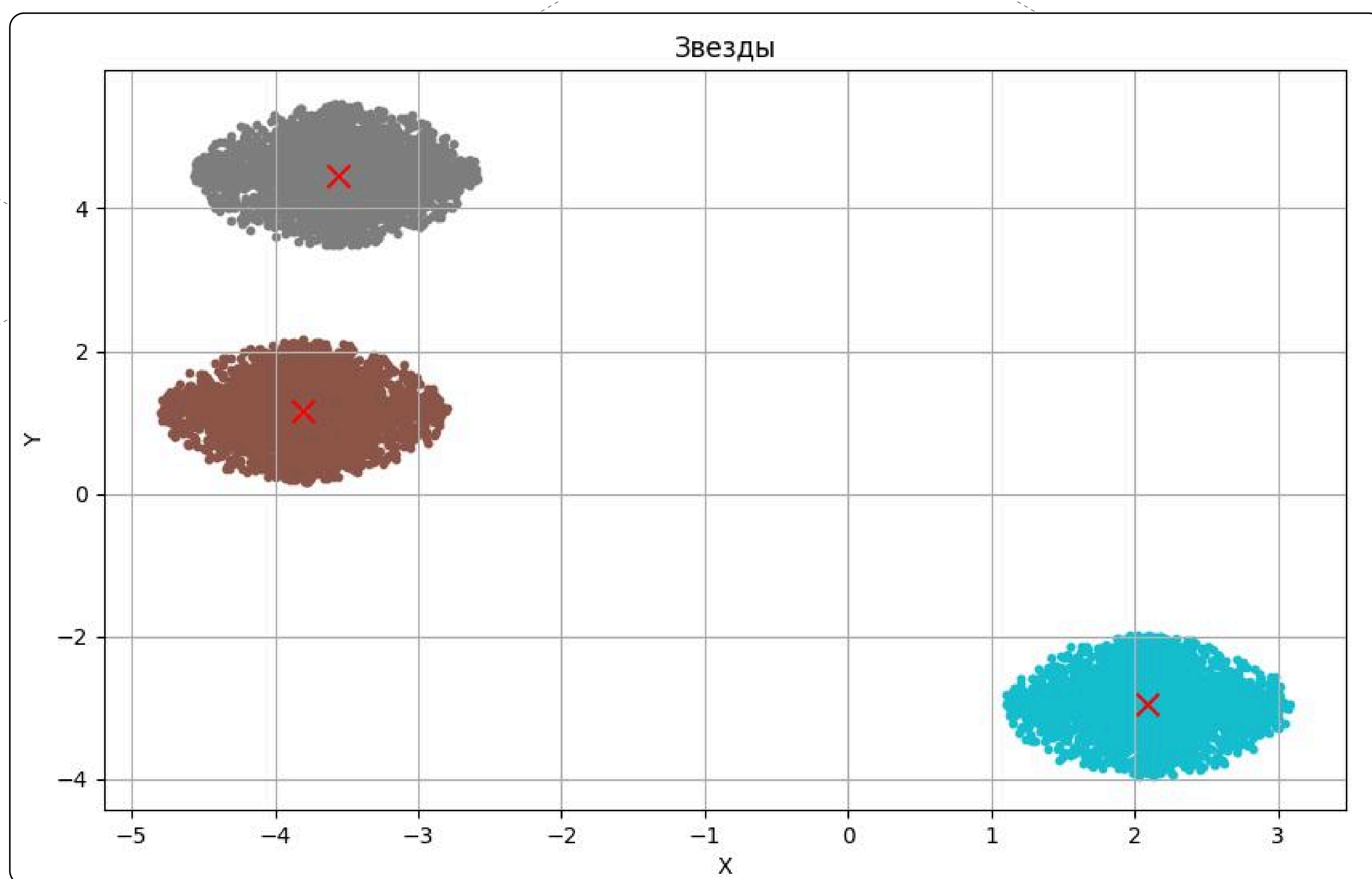
$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}.$$

P_{xy} — сумма абсцисс и ординат истинных центроидов кластеров, и P_H — сумма значений H для каждого кластера.

В ответе запишите четыре числа, в первой клетке целые части: $P_{xy} \times 10000$ и $P_H \times 10000$ для файла A , во второй клетке — аналогичные значения для файла B .

Возможные данные одного из файлов иллюстрированы графиком.





Задание №2

Ученый решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звезд — это набор звёзд (точек) на графике, лежащий внутри прямоугольника высотой H и шириной W . Каждая звезда обязательно принадлежит только одному из кластеров.

Истинный центр кластера, или центроид, — это одна из звёзд на графике, сумма евклидовых расстояний от которой до всех остальных звёзд кластера минимальна.

Для двух точек $A(x_1, y_1)$, $B(x_2, y_2)$ евклидово расстояние вычисляется по формуле:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

В файле А хранятся данные о звёздах двух кластеров, где $H=4$, $W=4$ для каждого кластера. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y , затем значение яркости звезды. Значения даны в условных единицах. Известно, что количество звёзд не превышает 1000.

В файле В хранятся данные о звёздах трёх кластеров, где $H=4$, $W=4$ для каждого кластера. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звёздах в файле В аналогична файлу А.

Для каждого файла найдите сумму абсцисс и ординат истинных центроидов кластеров, а затем параметр M — медианное значение яркости звезды в каждом кластере.

Медианное значение — это статистическая мера, которая представляет собой значение, делящее упорядоченный набор данных на две равные части. Если количество элементов в наборе нечётное, медианой будет средний элемент. Если количество элементов чётное, то медианой считается среднее арифметическое двух центральных элементов.



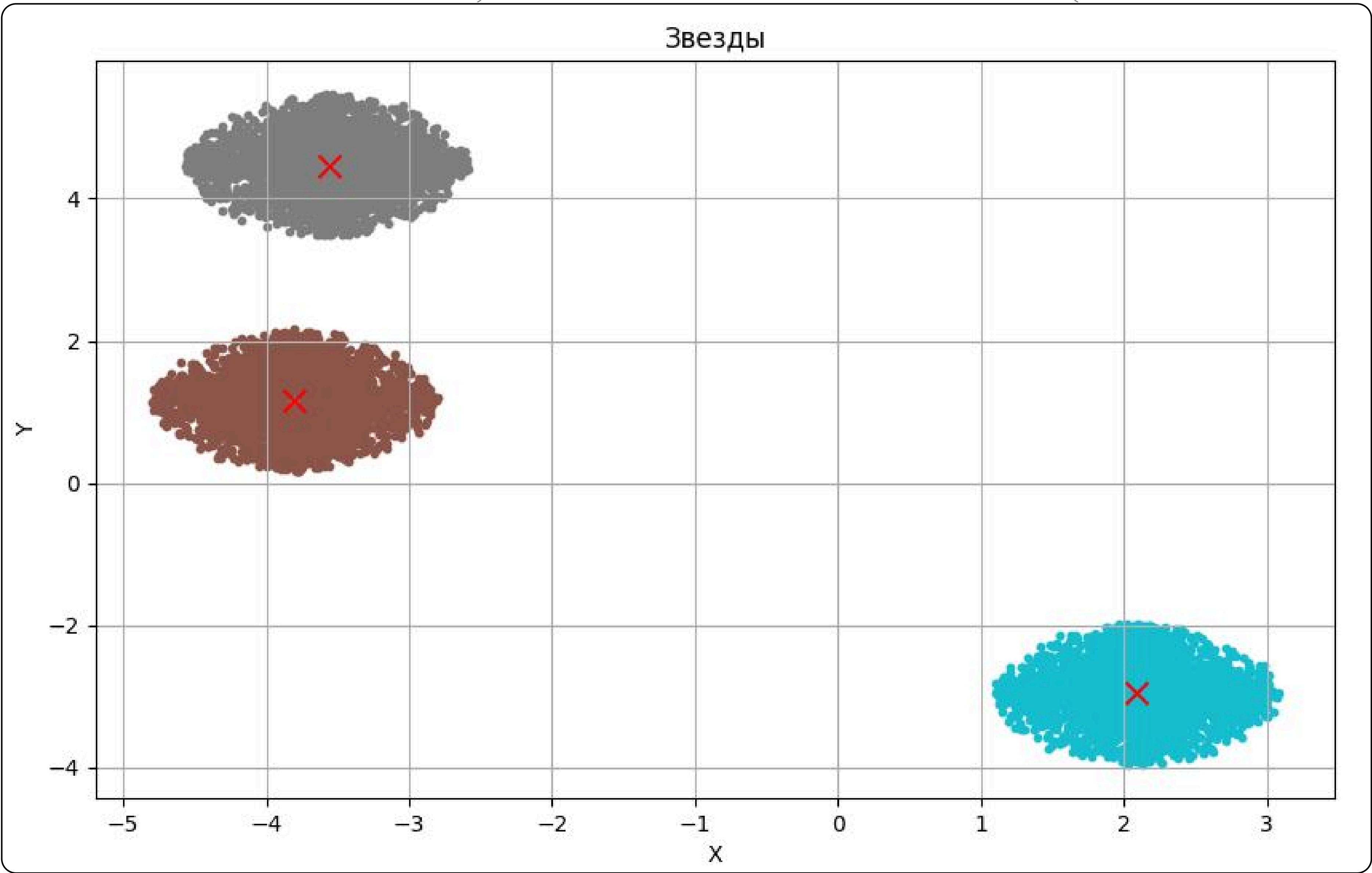
Пример:

- Для набора [1, 3, 5, 7, 9] медиана = 5 (средний элемент).
- Для набора [2, 4, 6, 8] медиана = $(4 + 6) / 2 = 5$.

P_{xy} — сумма абсцисс и ординат истинных центроидов кластеров, и P_M — сумма значений M для каждого кластера.

В ответе запишите четыре числа, в первой клетке целые части: $P_{xy} \times 10000$ и $P_n \times 10000$ для файла А, во второй клетке — аналогичные значения для файла В.

Возможные данные одного из файлов иллюстрированы графиком.



Ссылка на файлы 

ОТВЕТЫ

1. 6435 2387
10688 2607
2. 26309 10019
-48212 14839