

# Визуализация алгоритма кластеризации

## Задание №1

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – 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 расстояний от которой до всех остальных звёзд кластера минимальна. Под расстоянием понимается расстояние Евклида между двумя точками  $A(x_1, y_1)$  и  $B(x_2, y_2)$  на плоскости, которое 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  $H=3$ ,  $W=3$  для каждого кластера. В каждой строке записана информация о расположении на карте одной звезды: сначала координата  $x$ , затем координата  $y$ . Значения даны в условных единицах.

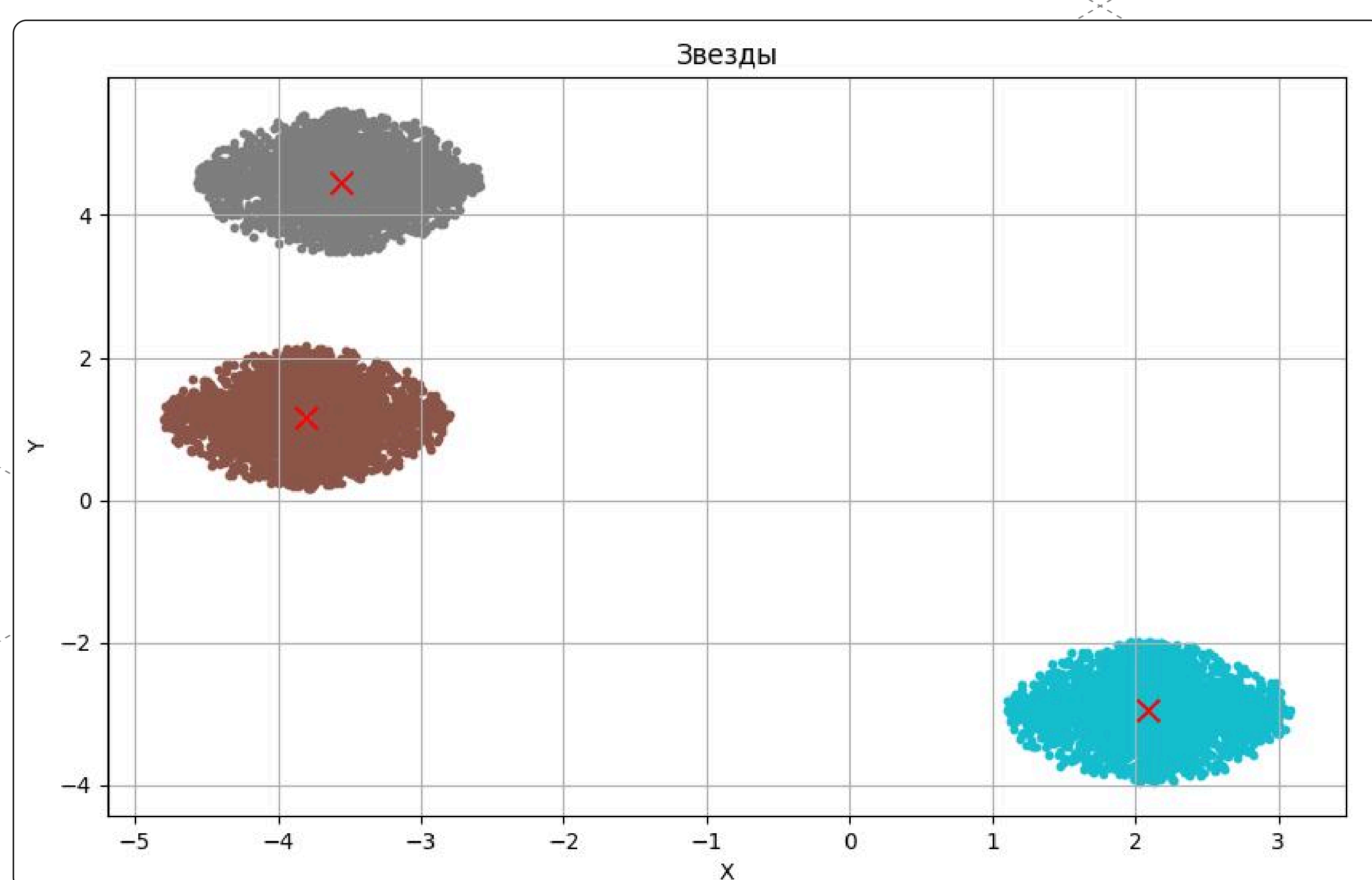
Известно, что количество звёзд не превышает 1000. В файле В хранятся данные о звёздах трёх кластеров, где  $H=3$ ,  $W=3$  для каждого кластера. Известно, что количество звёзд не превышает 11 000.

Структура хранения информации о звездах в файле В аналогична файлу А.

Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа:  $P_x$  – среднее арифметическое абсцисс центров кластеров, и  $P_y$  – среднее арифметическое ординат центров кластеров.

В ответе запишите четыре числа: в первой строке сначала целую часть произведения  $P_x \times 10\,000$ , затем целую часть произведения  $P_y \times 10\,000$  для файла А, во второй строке – аналогичные данные для файла В.

Возможные данные одного из файлов иллюстрированы графиком.





# Задание №2

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – 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 расстояний от которой до всех остальных звёзд кластера минимальна. Под расстоянием понимается расстояние Евклида между двумя точками  $A(x_1, y_1)$  и  $B(x_2, y_2)$  на плоскости, которое 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  $y$ . Значения даны в условных единицах.

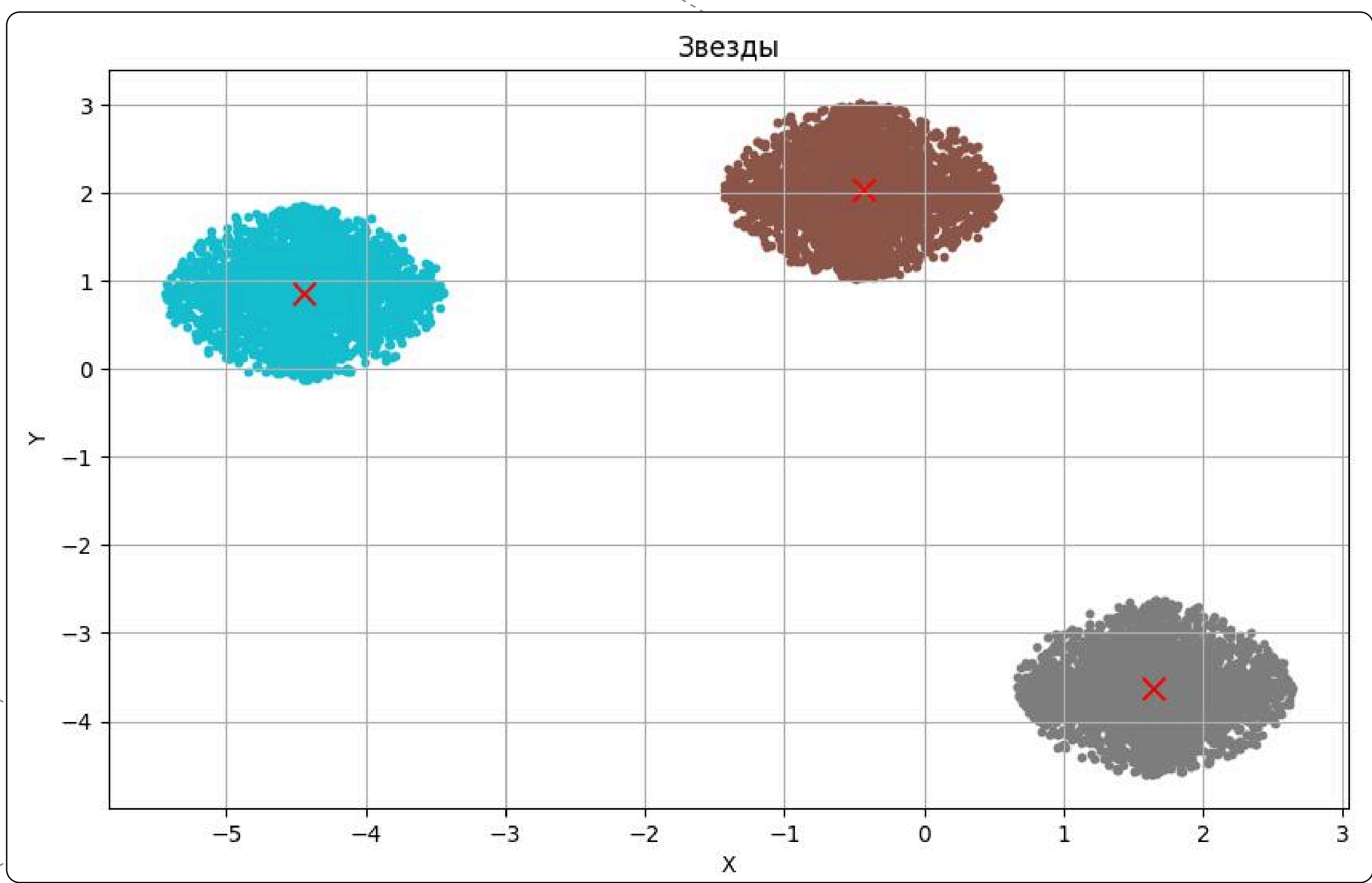
Известно, что количество звёзд не превышает 1000. В файле В хранятся данные о звёздах трёх кластеров, где  $H=4, W=4$  для каждого кластера. Известно, что количество звёзд не превышает 11 000.

Структура хранения информации о звездах в файле В аналогична файлу А.

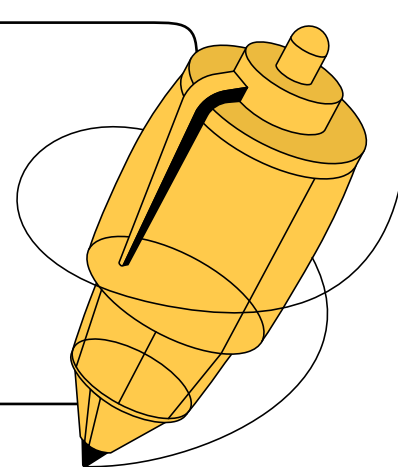
Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа:  $P_x$  – среднее арифметическое абсцисс центров кластеров, и  $P_y$  – среднее арифметическое ординат центров кластеров.

В ответе запишите четыре числа: в первой строке сначала целую часть произведения  $P_x \times 10\,000$ , затем целую часть произведения  $P_y \times 10\,000$  для файла А, во второй строке – аналогичные данные для файла В.

Возможные данные одного из файлов иллюстрированы графиком.



[Ссылка на файлы](#)



1. -50662-45182
2. -16035 18101
- 1955 947-746 5551

ОТВЕТЫ

# Заметки

