

DBSCAN. 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. Под евклидовым расстоянием понимается сумма модулей разностей координат двух точек. Для двух точек $A(x_1, y_1)$, $B(x_2, y_2)$ евклидово расстояние вычисляется по формуле:

В файле A 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.

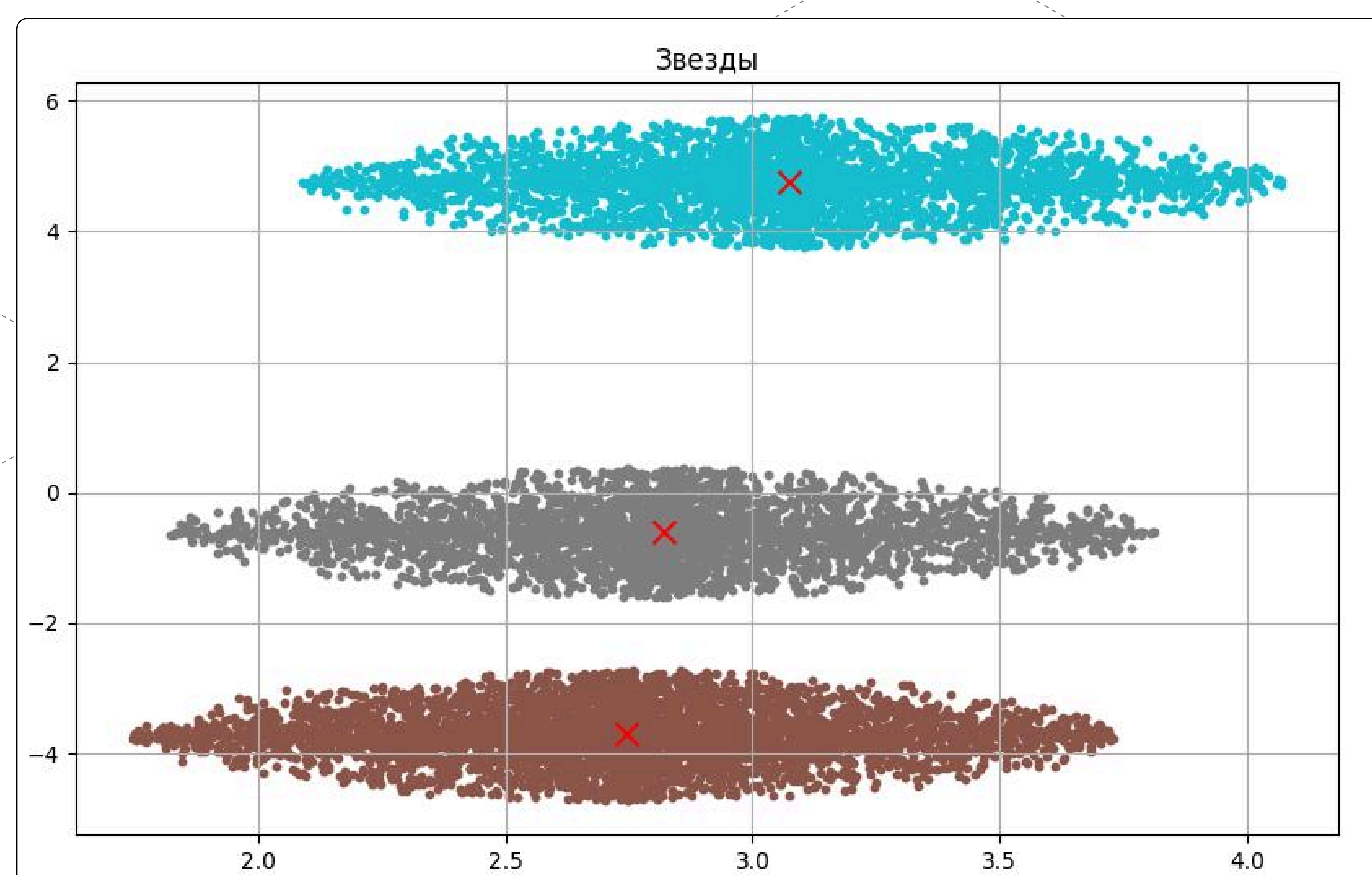
Известно, что количество звёзд не превышает 1000.

В файле B хранятся данные о звёздах трёх кластеров, где $H=3$, $W=3$ 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 определите координаты центра каждого кластера, используя евклидовое расстояние, 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 две строки через пробел: в первую строку целые части произведений: $P_x \times 10000$ и $P_y \times 10000$ для файла A , во вторую строку — 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.

Аномалиями считаются группы из 10 и менее точек, если каждая из них расположена более чем на половину условной единицы от ближайших точек в кластерах. Аномалии исключаются из расчётов.

Истинный центр кластера, или центроид, — это одна из звёзд на графике, сумма евклидовых расстояний от которой до всех остальных звёзд кластера минимальна. Под евклидовым расстоянием понимается сумма модулей разностей координат двух точек. 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 $H=3$, $W=3$ 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 10 000. Структура хранения информации о звёздах в файле В аналогична файлу А.

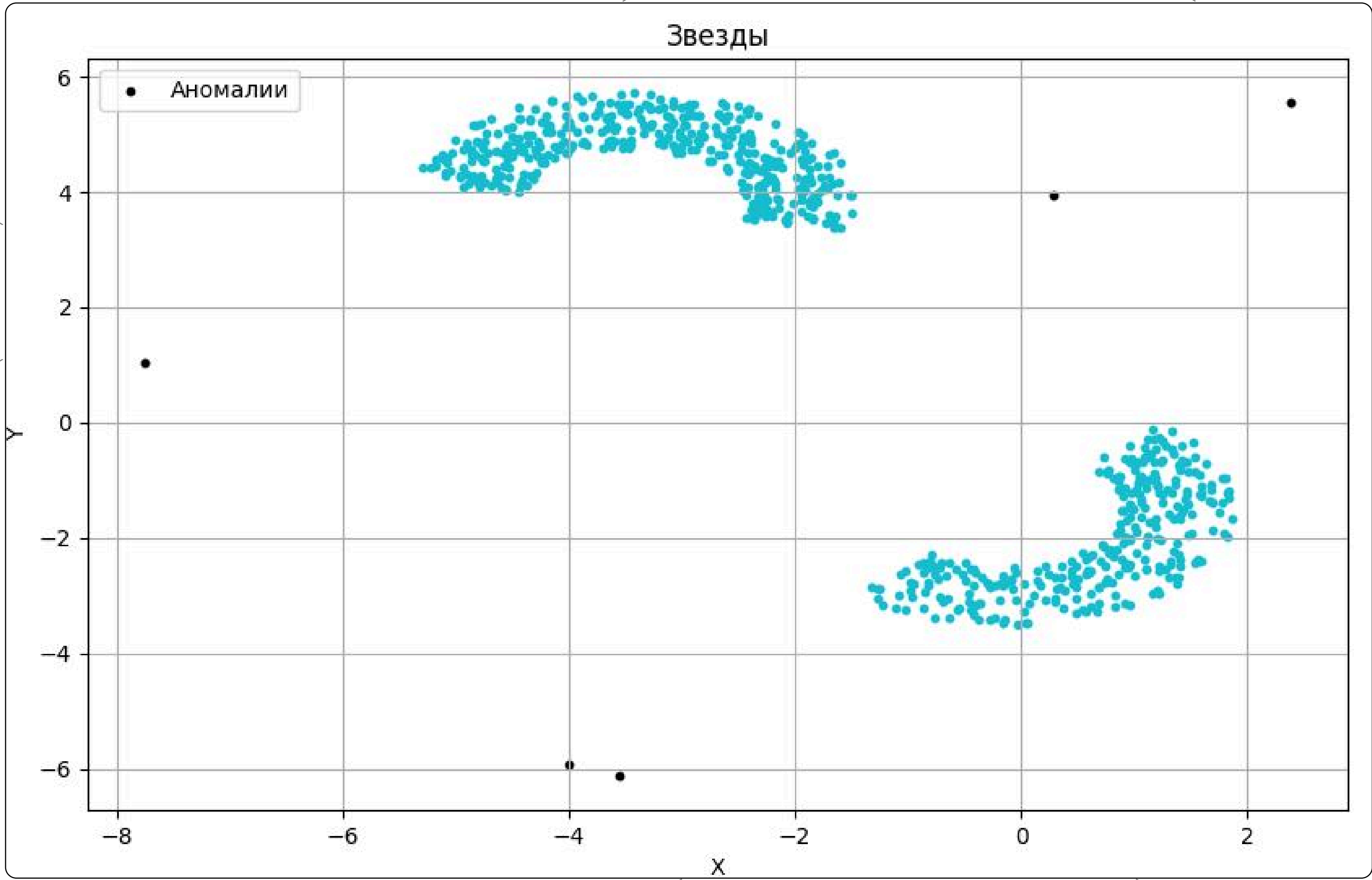
Для каждого файла определите координаты центра каждого кластера, используя евклидовое расстояние, затем вычислите два числа: P_x — среднее арифметическое абсцисс центров кластеров, и P_y — среднее арифметическое ординат центров кластеров, а P_s — сумма плотностей кластеров.

Плотность определяется как количество звёзд, делённое на площадь кластера.

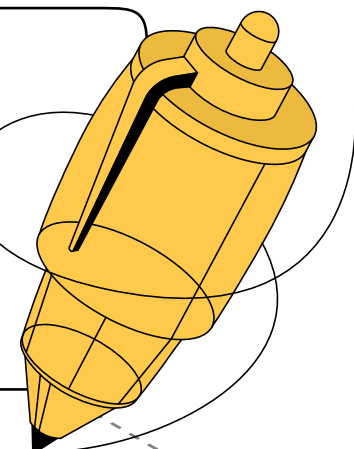
В ответе запишите четыре числа в две строки через пробел: в первую строку целые части произведений: $(P_x + P_y) \times 10000$ и $P_s \times 1000$ для файла А, во вторую строку — аналогичные значения для файла В.

Возможные данные одного из файлов иллюстрированы графиком.





Ссылка на файлы



- 1. 6446-50759
- 2881410788
- 2. -5698104888
- 17090904444

ОТВЕТЫ



Заметки

