

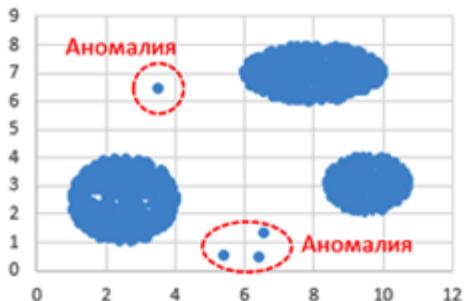
Домашнее задание 2. Задание 27 DB_Scan

№1 (7.12 (П) 1-2)

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – это набор не менее чем из 30 соседних звёзд (точек) на графике. Каждая звезда обязательно принадлежит только одному из кластеров. Центр кластера – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера минимальна. Расстояние между двумя точками $A(x_1, y_1)$ и $B(x_2, y_2)$ вычисляется по формуле:

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

Аномалиями назовём точки, находящиеся на расстоянии более одной условной единицы от точек кластеров. При расчётах аномалии учитывать не нужно. Даны два входных файла. В файле А хранятся данные о звёздах двух кластеров. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y (в условных единицах). Известно, что количество звёзд не превышает 1000. В файле Б хранятся данные о звёздах трёх кластеров. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А. Возможные данные одного из файлов иллюстрированы графиком. Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: R_x – среднее арифметическое абсцисс центров кластеров, и R_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $R_x \times 10 000$, затем целую часть произведения $R_y \times 10 000$ для файла А, во второй строке – аналогичные данные для файла Б.



Ответ:

19595 72247

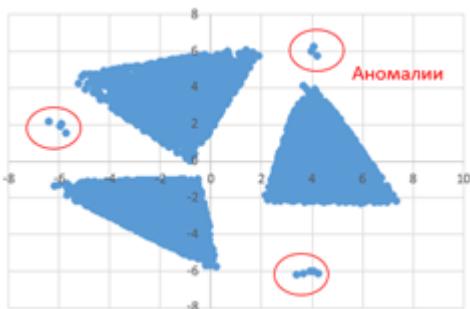
38012 50002

№2 (7.12 (П) 3-4)

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – это набор не менее чем из 30 соседних звёзд (точек) на графике. Каждая звезда обязательно принадлежит только одному из кластеров. Центр кластера, или центроид, – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера минимальна. Расстояние между двумя точками $A(x_1, y_1)$ и $B(x_2, y_2)$ вычисляется по формуле:

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

Аномалиями назовём точки, находящиеся на расстоянии более одной условной единицы от точек кластеров. При расчётах аномалии учитывать не нужно. Даны два входных файла. В файле А хранятся данные о звёздах двух кластеров. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y (в условных единицах). Известно, что количество звёзд не превышает 1000. В файле Б хранятся данные о звёздах трёх кластеров. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А. Возможные данные одного из файлов иллюстрированы графиком. Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: R_x – среднее арифметическое абсцисс центров кластеров, и R_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $R_x \times 100\ 000$, затем целую часть произведения $R_y \times 100\ 000$ для файла А, во второй строке – аналогичные данные для файла Б.



Ответ:

559397 625605

923413 613488

№3 (7.12 (П) 5-6)

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. Каждая звезда обязательно принадлежит только одному из кластеров.

Истинный край кластера – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера максимальна.

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:

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

Аномалиями назовём звёзды, находящиеся на расстоянии более одной условной единицы от звёзд кластеров. При расчётах аномалии учитывать не нужно.

В файле А хранятся данные о звёздах трёх кластеров. В каждой строке записана информация о расположении на карте одной звезды: сначала координата xx, затем координата yy. Значения даны в условных единицах. Известно, что количество звёзд не превышает 1000.

В файле В хранятся данные о звёздах пяти кластеров. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле В аналогична файлу А.

Для каждого файла определите координаты края каждого кластера, затем вычислите два числа: Тх – среднее арифметическое абсцисс краев кластеров, и Ту – среднее арифметическое ординат краев кластеров.

В ответе запишите четыре числа: в первой строке сначала целую часть произведения $Tx \times 10000$, затем целую часть произведения $Ty \times 10000$ для файла А, во второй строке – аналогичные данные для файла В.

Ответ:

13258 2656

-209434 474989

№4 (7.12 (П) 7-8)

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – это набор не менее чем из 30 соседних звёзд (точек) на графике. Каждая звезда обязательно принадлежит только одному из кластеров. Центр кластера, или центроид, – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера минимальна. Расстояние между двумя точками $A(x_1,y_1)$ и $B(x_2,y_2)$ вычисляется по формуле

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

Аномалиями назовём точки, находящиеся на расстоянии более одной условной единицы от точек кластеров. При расчётах аномалии учитывать не нужно. В файле А хранятся данные о звёздах двух кластеров. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y (в условных единицах). Известно, что количество звёзд не превышает 1000. В файле Б хранятся данные о звёздах трёх кластеров. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А. Возможные данные одного из файлов иллюстрированы

графиком.

Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: PxP – среднее арифметическое абсцисс центров кластеров, и PyP – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $Px \times 100000$, затем целую часть произведения $Py \times 100000$ для файла А, во второй строке – аналогичные

данные

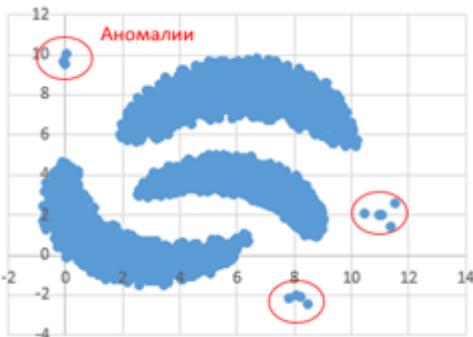
для

файла

Б.

Внимание! График приведён в иллюстративных целях для произвольных значений, не имеющих отношения к заданию.

Для выполнения задания используйте данные из прилагаемого файла.



Ответ:

358310 347120

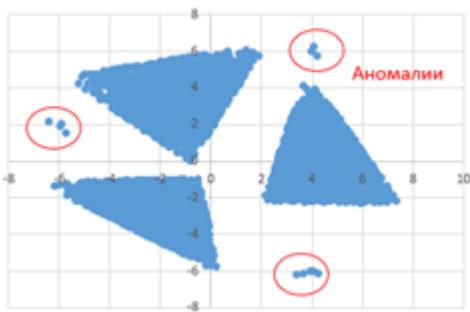
507819 522848

№5 (7.12 (П) 9-10)

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – это набор не менее чем из 30 соседних звёзд (точек) на графике. Каждая звезда обязательно принадлежит только одному из кластеров. Центр кластера, или центроид, – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера минимальна. Расстояние между двумя точками $A(x_1, y_1)$ и $B(x_2, y_2)$ вычисляется по формуле:

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

Аномалиями назовём точки, находящиеся на расстоянии более одной условной единицы от точек кластеров. При расчётах аномалии учитывать не нужно. Даны два входных файла. В файле А хранятся данные о звёздах двух кластеров. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y (в условных единицах). Известно, что количество звёзд не превышает 1000. В файле Б хранятся данные о звёздах трёх кластеров. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А. Возможные данные одного из файлов иллюстрированы графиком. Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: R_x – среднее арифметическое абсцисс центров кластеров, и R_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $R_x \times 100\ 000$, затем целую часть произведения $R_y \times 100\ 000$ для файла А, во второй строке – аналогичные данные для файла Б.



Ответ:

156151 194521

24156 32378