

Информатика. Звезды и кластеры. Часть 3.

Пример:

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба, а также по блеску звезды. Кластер звёзд – это набор звёзд (точек) на графике, лежащий внутри прямоугольника высотой H и шириной W . Каждая звезда, подходящая под заданный уровень блеска, обязательно принадлежит только одному из кластеров. Остальные звёзды не относятся к рассматриваемым кластерам.

Истинный центр кластера, или центроид, – это одна из звёзд кластера, 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 $H = 5$, $W = 5$, для каждого кластера, а звёзды обладают блеском не более 20 условных единиц. В каждой строке записана информация о расположении на карте одной звезды, а также об её уровне блеска: сначала координата x , затем координата y и наконец уровень блеска m . Значения даны в условных единицах, которые представлены вещественными числами. Известно, что количество звёзд не превышает 2000.

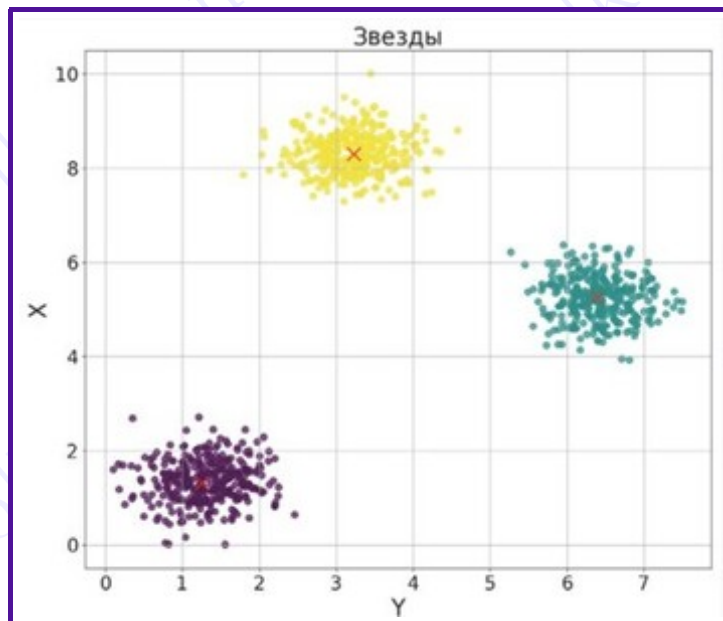
В файле Б хранятся данные о звёздах **четырёх** кластеров, где $H = 5$, $W = 5$ для каждого кластера, а звёзды обладают блеском не более 20 условных единиц. Известно, что количество звёзд не превышает 20 000. Структура хранения информации о звёздах в файле Б аналогична файлу А.

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 через пробел: сначала целую часть произведения $P_x \times 10000$ для файла А и $P_y \times 10000$ для файла А, далее целую часть произведения $P_x \times 10000$ для файла Б и $P_y \times 10000$ для файла Б.

Возможные данные одного из файлов иллюстрированы графиком.

Внимание! График приведён в иллюстративных целях для произвольных значений, не имеющих отношения к заданию. Для выполнения задания используйте данные из прилагаемого файла.



Решение:

У нас есть 4 кластера:

1 — 1...5, где значения от 1 до 5 — блеск

2 — 6..10

3 — 11...15

4 — 16...20

```

f = open('27__4x6s7.txt')
a = [[] for i in range(4)] # Так как у нас 4 кластера
for i in f: # Берём данные из нашего файла
    x, y, r = map(float, i.split()) # Читаем всё, как вещественные числа
    if r in range(1, 6): # Раскидываем по кластерам наши звёзды
        a[0].append([x, y])
    if r in range(6, 11):
        a[1].append([x, y])
    if r in range(11, 16):
        a[2].append([x, y])
    if r in range(16, 21):
        a[3].append([x, y])
# Проходимся по каждому кластеру, перебираем звёзды и находим центр
# Нам нужно посчитать сумму абсцисс и сумму ординат наших центроидов
sx = sy = 0
for k in range(4):
    mn = 10000000050000
    for j in a[k]:
        # Находим сумму длин расстояний для каждой звезды внутри кластера
        s = 0
        x1, y1 = j
        for i in a[k]: # Проходимся по кластеру, берём координаты
            # каждой звезды, считаем расстояние до этой звезды
            x2, y2 = i
            s += ((x2 - x1) ** 2 + (y2 - y1) ** 2) ** 0.5
        if s < mn: # Ищем минимальное расстояние по каждому кластеру
            mn = s
            tx, ty = x1, y1
    sx += tx
    sy += ty
print(int(sx/4 * 10000), int(sy/4 * 10000))

```

Ответ: 26074 25360