

## Информатика. Задание на кластеризацию. Повторение материала.

### Задача

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 минимальна. Под расстоянием понимается расстояние Евклида между двумя точками  $A(x_1, y_1, z_1)$  и  $B(x_2, y_2, z_2)$  на плоскости, которое вычисляется по формуле:

$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

### Входные данные

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  $y$ , затем координата  $z$ . Значения даны в условных единицах, которые представлены вещественными числами. Известно, что количество звёзд не превышает 1000.

В файле Б хранятся данные о звёздах трех кластеров, где  $H = 3$ ,  $W = 3$  для каждого кластера. Известно, что количество звёзд не превышает 10000. Структура хранения информации о звездах в файле Б аналогична файлу А.

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

### Выходные данные

В ответе запишите четыре числа через пробел: сначала целую часть произведения  $P_x \cdot 1000$  для файла А,  $P_y \cdot 1000$  для файла А,  $P_z \cdot 1000$  для файла А, далее целую часть произведения  $P_x \cdot 1000$  для файла Б,  $P_y \cdot 1000$  для файла Б и  $P_z \cdot 1000$  для файла Б.

### Решение для файла А:

Для начала визуально оценим данные в условии кластеры. Для этого откроем предложенные файлы в *Excel*, перейдем в раздел «Вставка → Диаграммы → Точечная». Таким способом можно построим по отдельности двухмерные проекции кластеров на оси  $XoY$ ,  $YoZ$  и  $XoZ$ .

Диаграммы для файла А имеют вид:

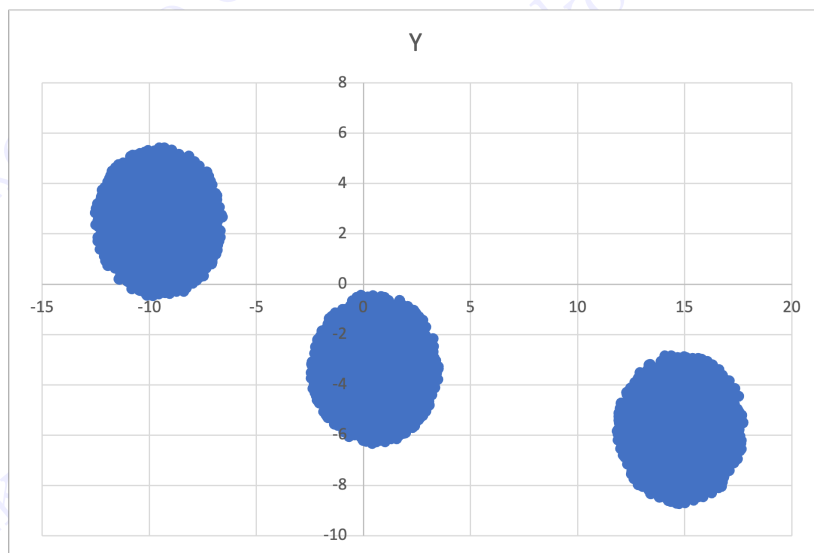


Диаграмма проекций на оси  $XoY$

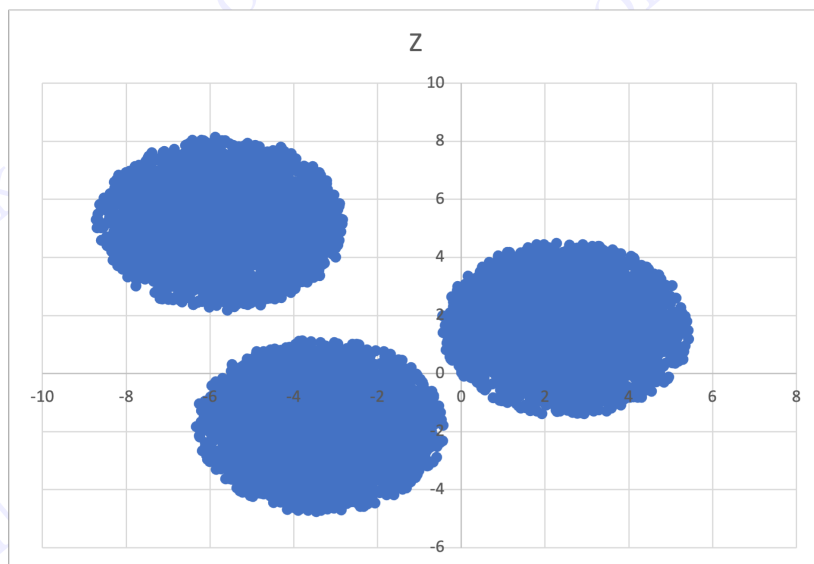


Диаграмма проекций на оси  $YoZ$

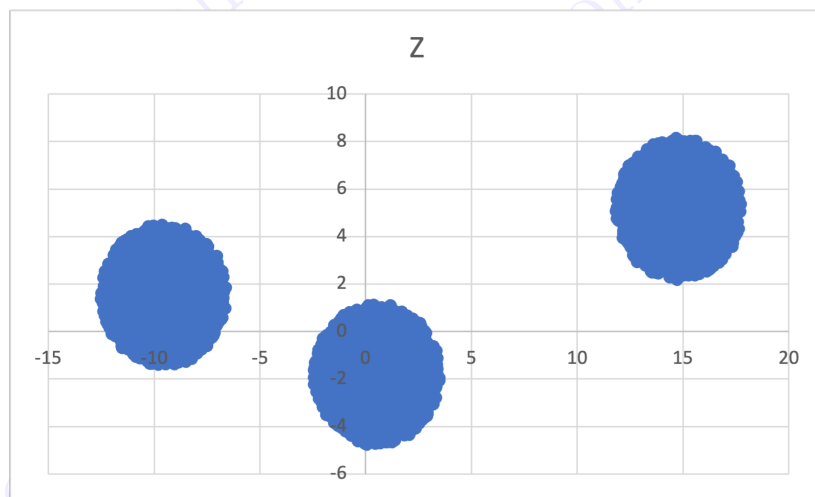


Диаграмма проекций на оси  $XoZ$

Исходя из диаграмм проекций, становится ясно, что звезды первого кластера имеют координаты по оси  $X$  меньше  $-5$ , звезды второго кластера имеют координаты по оси  $X$  больше  $-5$ , но меньше  $5$ , а оставшиеся звезды – звезды третьего кластера.

---

```

from math import
f=open('27_3.txt')
s=f.readline()
a=[list(map(float, i.replace(',', '.').split())) for i in f]
# Считываем строки как звезды
clusters = [[], [], []] # Создаем список для хранения кластеров
for i in a: # Распределяем звезды по кластерам в соответствии с диаграммами
    x, y, z = i
    if x<-5:
        clusters[0].append(i)
    elif x<5:
        clusters[1].append(i)
    else:
        clusters[2].append(i)
sum_x = sum_y = sum_z = 0
# Переменные для хранения сумм координат центроидов
for i in clusters: # Проходимся циклом по кластерам
    mn = 100000500000 # Переменная для хранения минимального расстояния
    for k in i:
        s = 0
        star = k # Звезда - возможный центроид
        for j in i:
            s += dist(star, j)
            # Суммируем расстояние от звезды star до остальных звезд
        if s < mn: # Если расстояние до star
            # Оказалось меньше расстояний до других звезд
            mn = s # Перезаписываем минимальное расстояние
            t = star # Сохраняем новую звезду как центроид
    sum_x += star[0] # Сумма абсцисс центроидов
    sum_y += star[1] # Сумма ординат центроидов
    sum_z += star[2] # Сумма аппликат центроидов
print(int(sum_x/3*1000), int(sum_y/3*1000), int(sum_z/3*1000))

```

---

Ответ: 2658 -2722 2426