

Информатика. Звёзды и кластеры. Часть 1.

1 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 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:

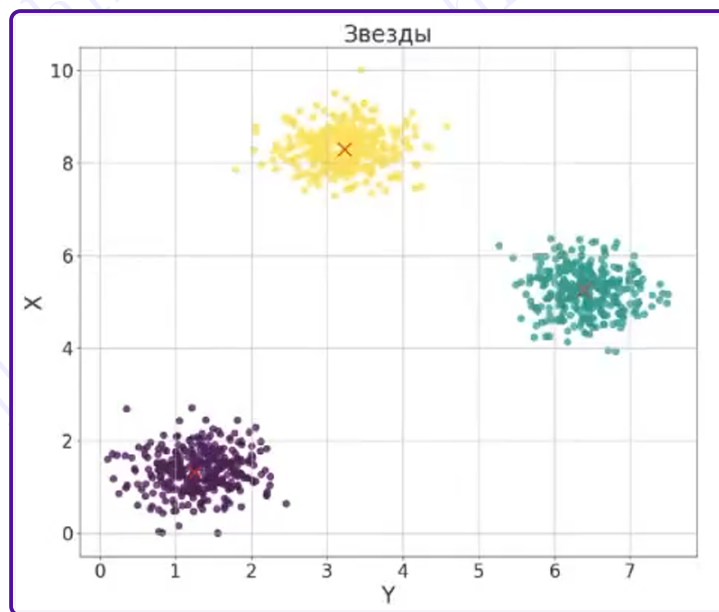
$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

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 Б 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 звездах в файле Б аналогична файлу А.

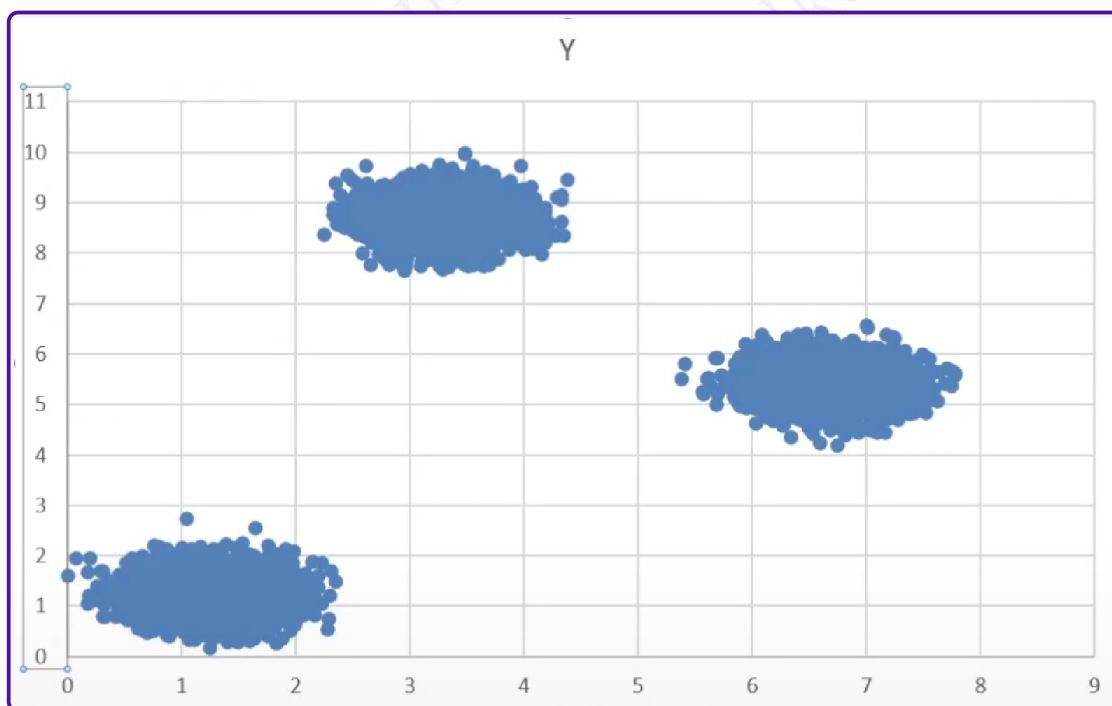
Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: P_x – среднее арифметическое абсцисс центров кластеров, и P_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $P_x \times 10\,000$, затем целую часть произведения $P_y \times 10\,000$ для файла А, во второй строке – аналогичные данные для файла 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:

Открываем файл Excel, выделяем полностью два столбца X и Y , переходим в раздел «Вставка», выбираем точечную диаграмму. В параметрах указываем деления – 1. Получили кластеры.



В соответствии с диаграммой получаем, что если координата меньше трех по оси X и меньше трех по оси Y , это левый нижний кластер. Если $Y > 7$, то это верхний средний кластер. Если $X > 5$, то это левый нижний кластер.

```

f = open('demo_2025_27_B.txt') # Открываем файл
s = f.readline() # Считываем первую строку с названием столбцов
a = [[] for i in range(3)] # Создаем массив для деления на кластеры
for i in range(10000):
    # Считываем строку и меняем в числе ',' на '.'
    s = f.readline().replace(',', '.')
    # Делим строку на числа float и записываем их в массив
    xy = list(map(float, s.split()))
    if (xy[0] < 3) and (xy[1] < 3): # Если координаты 'x' < 3 и 'y' < 3
        a[0].append(xy) # Записываем координаты звезды в первый кластер
    elif xy[1] > 7: # Если 'y' > 7
        a[1].append(xy) # Записываем координаты звезды во второй кластер
    else:
        a[2].append(xy) # Записываем координаты звезды в третий кластер

# Поиск центроидов
sx = sy = 0 # Искомые средн. арифметические
for k in range(3): # Для каждого кластера
    mn = 1000000500000000000 # Минимальное расстояние
    for j in range(len(a[k])): # Для каждой звезды кластера
        ls = 0 # Длина расстояния
        x1, y1 = a[k][j] # Координаты звезды, от которой ищем расстояние
        for i in range(len(a[k])):
            x2, y2 = a[k][i] # Координаты звезды, до которой ищем расстояние
            ls += ((x2-x1)**2 + (y2-y1)**2)**0.5
        if ls < mn: # Если расстояние меньше минимума
            mn = ls # Обновляем mn
            t = a[k][j] # Запоминаем координаты предполагаемого центроида
sx += t[0] # Добавляем координату 'x' найденного центроида
sy += t[1] # Добавляем координату 'y' найденного центроида
print(int((sx/3)*10000), int((sy/3)*10000))

```

Ответ: 37522 51277