

Информатика. Звезды и кластеры. Часть 2.

1 Разбор заданий с вебинара

1.1 Пример 1

Ученый решил провести кластеризацию некоторого множества звезд по их расположению на карте звездного неба. Кластер звезд – это набор звезд (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 звезд на графике, сумма расстояний от которой до всех остальных звезд 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:

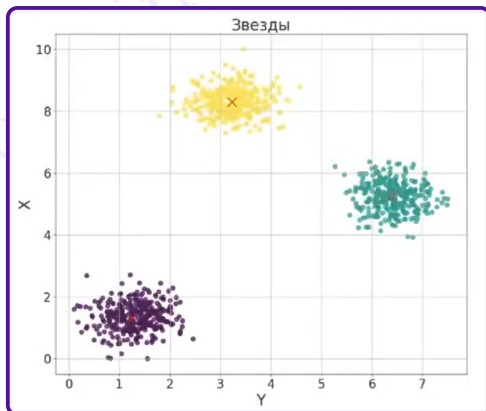
$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

В файле А хранятся данные о звездах **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 звезд не превышает 1000.

В файле Б хранятся данные о звездах **трех** кластеров, где $H = 3$, $W = 3$ для каждого кластера. Известно, что количество звезд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А.

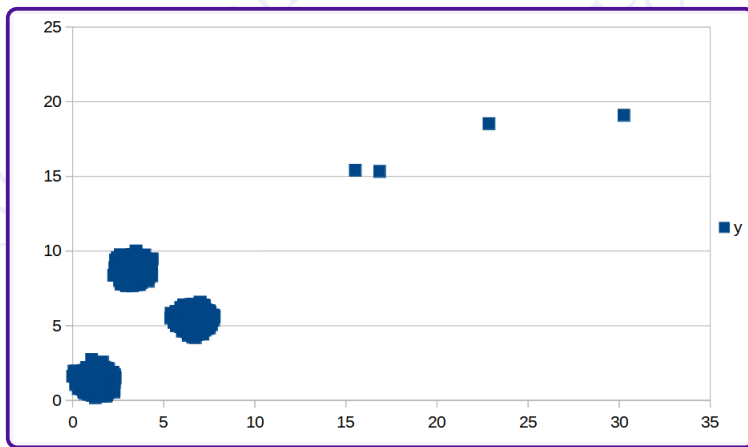
Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: P_x – среднее арифметическое абсцисс центров кластеров, и P_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $P_x \times 10\,000$,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



Исходя из диаграммы, рассмотрим три случая:

1) Если координата $X < 3$ и координата $Y < 4$, то точка (X, Y) принадлежит самому нижнему кластеру.

2) Если координата X находится в отрезке $[2; 5]$, а координата Y – в отрезке $[7; 11]$, то точка (X, Y) принадлежит самому верхнему кластеру.

3) Если же координата X находится в отрезке $[5; 8]$, а координата Y – в отрезке $[3; 7]$, то точка (X, Y) принадлежит кластеру посередине. Аномальные точки таким образом отсеиваются, так как не попадают ни под одно условие.

```

f = open('27.txt') # Открываем файл '27.txt' для чтения
s = f.readline() # Читаем первую строку
a = [[], [], []] # Создаем список для хранения координат звезд кластеров
for i in f: # Обрабатываем оставшиеся строки файла
    x, y = i.replace(',', '.').split()
    # Проверяем, содержат ли x и y только цифровые символы (включая точку)
    if (x+y).replace('.', '').isdigit():
        x = float(x) # Преобразуем x и y в числа с плавающей запятой
        y = float(y)
        if x < 3 and y < 4: # Кластер 1
            a[0].append([x, y])
        if (2 <= x <= 5) and (7 <= y <= 11): # Кластер 2
            a[1].append([x, y])
        if (5 <= x <= 8) and (3 <= y <= 7): # Кластер 3
            a[2].append([x, y])
sx = sy = 0 # Инициализируем переменные для накопления координат центроидов
for k in range(3):
    mn = 100000000050000000000
    for j in range(len(a[0])): # Для каждой звезды в текущем кластере
        x1, y1 = a[0][j] # Получаем координаты звезды
        s = 0 # Переменная для суммы расстояний до всех остальных звезд
        for i in a[0]: # Считаем расстояния до всех звезд в том же кластере
            x2, y2 = i
            s += ((x2-x1)**2 + (y2-y1)**2)**0.5
        if s < mn:
            mn = s
            tx = x1
            ty = y1
    sx += tx # Добавляем координаты найденного центра к общим координатам
    sy += ty
print(int(sx/3 * 10000), int(sy/3 * 10000))

```

Ответ: 37522 51277