Задание 27

Уровень сложности: ПростаяПростая

Автор: Яндекс УчебникЯндекс Учебник

Тема: Анализ данных(Анализ данных)

Фрагмент звёздного неба спроецирован на плоскость с декартовой системой координат, звёзды изображаются точками. Учёный решил провести кластеризацию этих точек, то есть разбить их множество на несколько непересекающихся подмножеств (кластеров). Точки каждого подмножества должны лежать внутри непересекающихся квадратов со стороной длиной H. Стороны квадратов не обязательно параллельны координатным осям. Гарантируется, что такое разбиение точно есть, и оно единственно для заданных размеров квадрата.

Поскольку в круглый объектив телескопа можно увидеть только область, ограниченную радиусом R = 2,1, учёный для удобства решил выбрать в каждом кластере одну звезду как центр наведения объектива телескопа (точку наведения). Основной параметр для выбора такой точки — максимально возможное количество звёзд, которые попадают при этом в поле зрения объектива (без учета звёзд на границе видимости). Если этот параметр одинаков для нескольких звёзд, среди них выбирается та, которая ближе всего находится к началу координат плоскости.

Расстояние между двумя точками на плоскости A(x1,y1)*A*(*x*1​,*y*1​) и B(x2,y2)*B*(*x*2​,*y*2​) вычисляется по формуле:

d(A, B) = (x2−x1)2+(y2−y1)2(*x*2​−*x*1​)2+(*y*2​−*y*1​)2​

В файле А хранятся координаты точек двух кластеров, где H = 4,2 для каждого кластера. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x, затем координата y. Известно, что количество точек не превышает 1000.

В файле Б хранятся координаты точек трёх кластеров, где H = 4,5 для каждого кластера. Известно, что количество точек не превышает 10000. Структура хранения информации в файле Б аналогична файлу А.

Для каждого файла определите координаты точки наведения каждого кластера, затем вычислите два числа: Px*Px*​ — среднее арифметическое абсцисс точек наведения кластеров, и Py*Py*​ — среднее арифметическое ординат точек наведения кластеров.

В ответе запишите четыре числа: в первой строке целую часть произведения ∣Px∣×10000∣*Px*​∣×10000, затем целую часть произведения ∣Py∣×10000∣*Py*​∣×10000 для файла A, во второй строке — аналогичные данные для файла Б.

График иллюстрирует возможные данные в файле A. **График приведён для иллюстративных целей, значения могут не соответствовать данным задачи.**

**Чтобы выполнить задание, используйте данные из прилагаемых файлов.**

<https://education.yandex.ru/ege/task/6106ffcb-0b89-43be-bb9e-e8e706c5bbe2>

Задание 27

<https://education.yandex.ru/ege/task/59c0d130-9a9c-4ec6-8847-c8ace888620a>

Уровень сложности: СредняяСредняя

Автор: Яндекс УчебникЯндекс Учебник

Тема: Анализ данных(Анализ данных)

Учёный решил исследовать карту звёздного неба. Небесные тела формируют созвездия — группы точек, которые задают контуры звёзд с пятью лучами разного размера. Каждое небесное тело принадлежит только одному созвездию.

В обоих файлах хранятся данные о небесных телах, формирующих множество созвездий. В каждой строке записана информация о расположении одного небесного тела: сначала координата x*x*, затем координата y*y*. Значения даны в условных единицах.

В **файле A**:

* общее количество небесных тел не превышает 10 00010000
* для дальнейших вычислений необходимо выбрать **два крупнейших созвездия** (по занимаемой площади)

В **файле Б**:

* общее количество небесных тел не превышает 100 000100000
* для дальнейших вычислений необходимо выбрать **три крупнейших созвездия** (по занимаемой площади)

**Задача:**

1. Для каждого из выбранных созвездий выполните следующие шаги:
   * найдите **среднее арифметическое расстояние** между всеми парами точек созвездия
   * найдите **максимальное расстояние** между любыми двумя точками в созвездии
   * рассчитайте **плотность созвездия** как  
     D=среднее расстояниемаксимальное расстояние*D*=максимальное расстояниесреднее расстояние​
   * под расстоянием понимается расстояние Евклида между двумя точками A(x1,y1)*A*(*x*1​,*y*1​) и B(x2,y2)*B*(*x*2​,*y*2​) на плоскости, которое вычисляется по формуле: d(A,B)=(x2−x1)2+(y2−y1)2*d*(*A*,*B*)=(*x*2​−*x*1​)2+(*y*2​−*y*1​)2​
2. Для каждого файла объедините результаты для выбранных созвездий следующим образом:
   * найдите **минимальную плотность** среди всех выбранных созвездий, обозначьте её как Pmin*P*min​
   * найдите **среднюю арифметическую плотность** всех выбранных созвездий, обозначьте её как Pavg*P*avg​

В ответе запишите четыре числа:

* в первой строке: целую часть произведения Pmin×10 000*P*min​×10000) и Pavg×10 000*P*avg​×10000 для файла A
* во второй строке: аналогичные данные для файла Б

График иллюстрирует возможные данные в файле A. **График приведён для иллюстративных целей, значения могут не соответствовать данным задачи.**

**Для выполнения задания используйте данные из прилагаемых файлов.**

Задание 27

Уровень сложности: СредняяСредняя

Автор: Яндекс УчебникЯндекс Учебник

Тема: Анализ данных(Анализ данных)

В лаборатории проводится серия экспериментов. Результат каждого измерения в эксперименте представляется в виде пары чисел. Для анализа результатов эта пара рассматривается как координаты точки на плоскости, и строится график, где отмечаются точки, соответствующие всем измерениям. По ошибке данные нескольких экспериментов были записаны в один файл.

Известно, что измерение относится к эксперименту, если расстояние до любой другой ближайшей точки не превышает 5.5 условных единиц.

Расстояние между двумя точками A(x1,y1)*A*(*x*1​,*y*1​) и B(x2,y2)*B*(*x*2​,*y*2​) вычисляется по формуле: d(A,B)=(x2−x1)2+(y2−y1)2.*d*(*A*,*B*)=(*x*2​−*x*1​)2+(*y*2​−*y*1​)2​.

Необходимо найти среди всех измерений те результаты экспериментов, у которых есть сильная корреляционная связь.

Две группы экспериментальных данных считаются сильно коррелирующими, если модуль коэффициента корреляции Пирсона между их результатами находится в диапазоне от 0,9 до 1. Коэффициент корреляции Пирсона (r*r*) для двух наборов данных (X*X* и Y*Y*) вычисляется по формуле:

r=∑i=1n(xi−x‾)(yi−y‾)∑i=1n(xi−x‾)2∑i=1n(yi−y‾)2*r*=∑*i*=1*n*​(*xi*​−*x*)2∑*i*=1*n*​(*yi*​−*y*​)2​∑*i*=1*n*​(*xi*​−*x*)(*yi*​−*y*​)​,

где x‾*x* и y‾*y*​ — среднеарифметические значения параметров X*X* и Y*Y* соответственно.

Графически сильная корреляционная связь может выглядеть так:

https://education.yandex.ru/ege/task/e6f51078-7ec7-4847-ad8d-3f6efa90ffe1