Описание тестовых данных:

Изначально происходит генерация данных. Реализованы 3 типа генераций: два кластера заведомо не вложенных (чтобы провести k-means), два кластера заведомо вложены (чтобы провести иерархическую кластеризацию), два пересекающихся кластера, центры которых не лежат в выпуклой оболочке других (пересечение двух колец).

Описание алгоритма:

1. Первоначально мы генерируем данных и делим их на несколько частей: разбиваем все данные на 2 массива (одна половина нужна для первого прогона, вторая для до обучения). Второй массив оставляем до лучших времен, а из первого массива делаем выборку (примерно 100 часть всех данных), которая будет нужна для иерархической кластеризации.
2. На полученной выборке проводим иерархическую кластеризацию для получения центров будущих кластеров и для определения вложенности кластеров.
3. Если кластеры вложены, значит запускаем иерархическую кластеризацию на всех данных (из первого массив), если же нет запускаем k-means.

3\*. Вложенность проверяется через выпуклую оболочку (построенную на каждом кластере) и трассировку луча (из центра кластера).

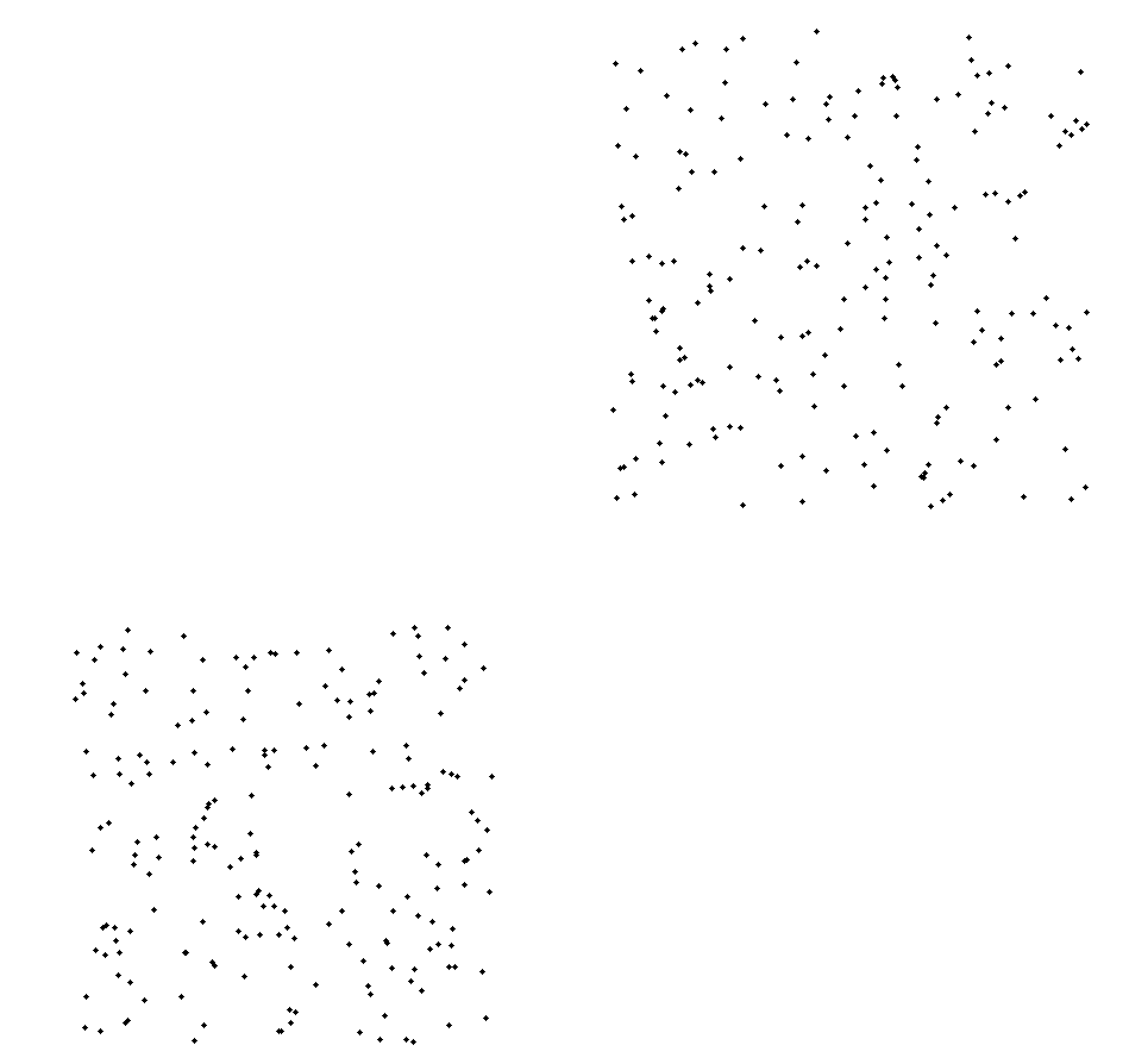
1. Затем запускается проверка k-средних, чтобы понять качественно ли выполнилась кластеризация
2. В самом конце запускается процесс до обучения: часть “старых” данных отбрасывается, а на оставшихся данных и добавленных (из первого массива) “новых” проводится кластеризация по тому же методу что был на первом прогоне.

Естественно, проверка на k-средних имела значение только для последней генерации данных, так как только там мы не знаем как поведет себя программа.

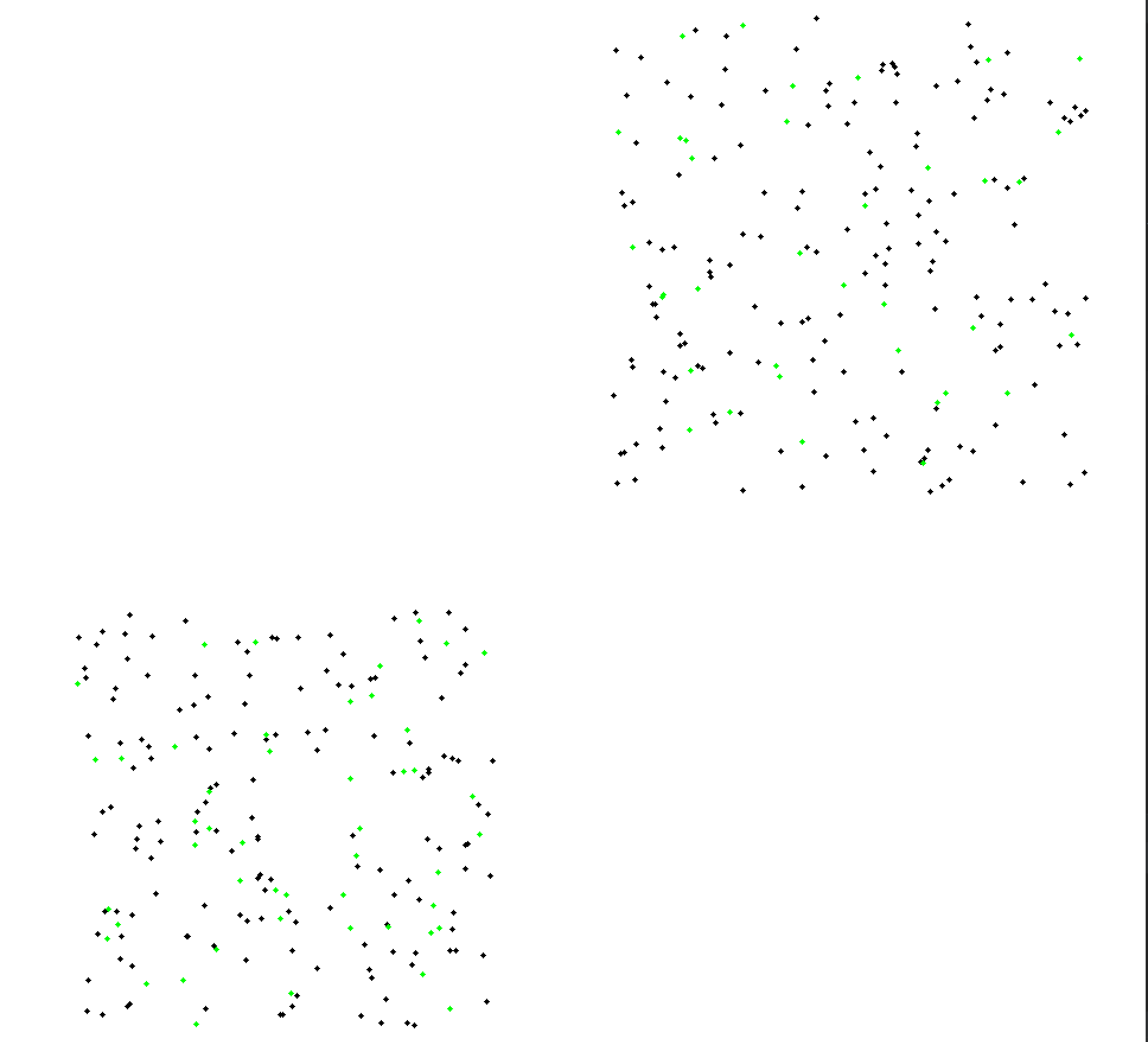
Результаты:

(Невложенные кластеры)

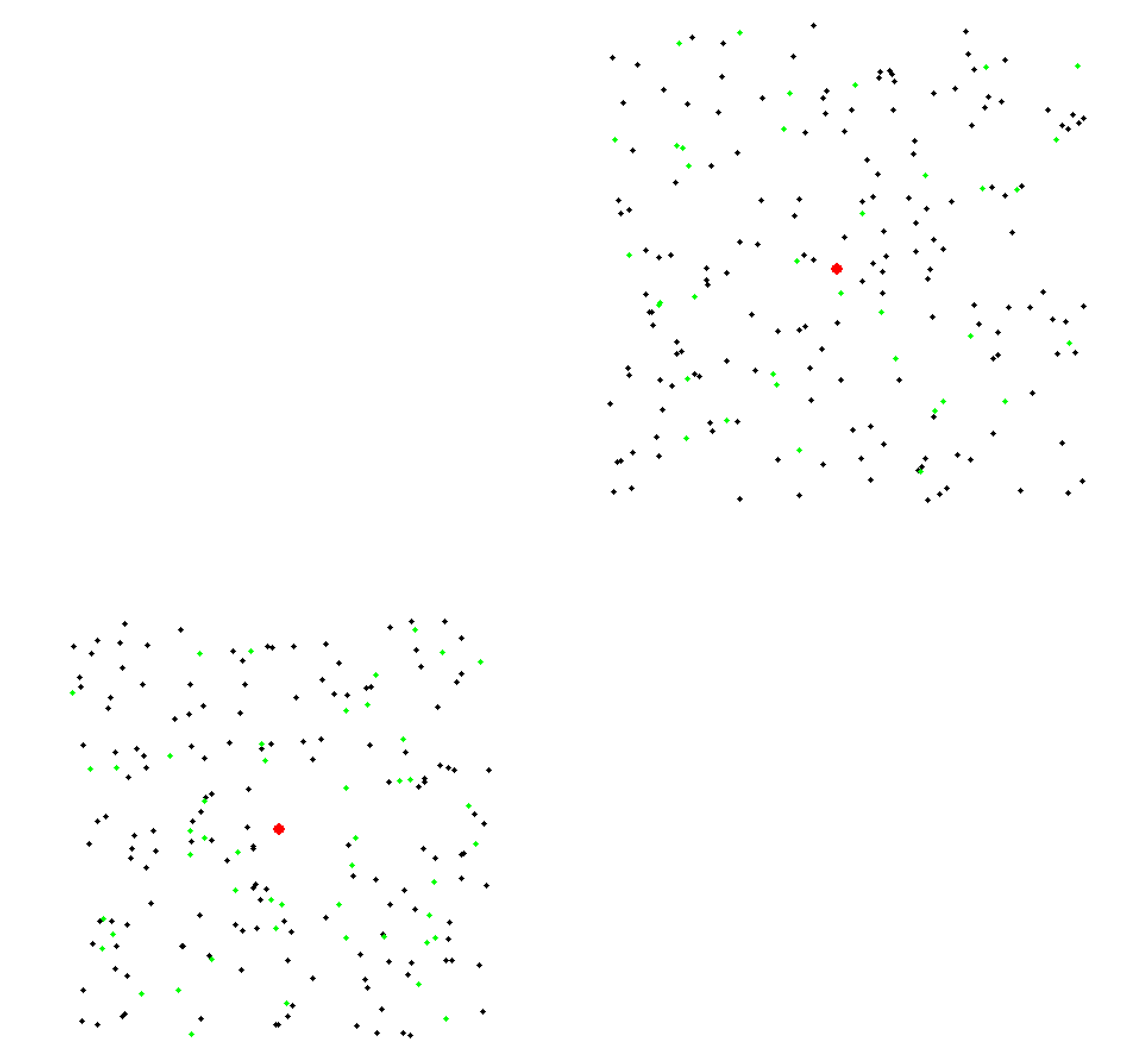
1. Генерация данных



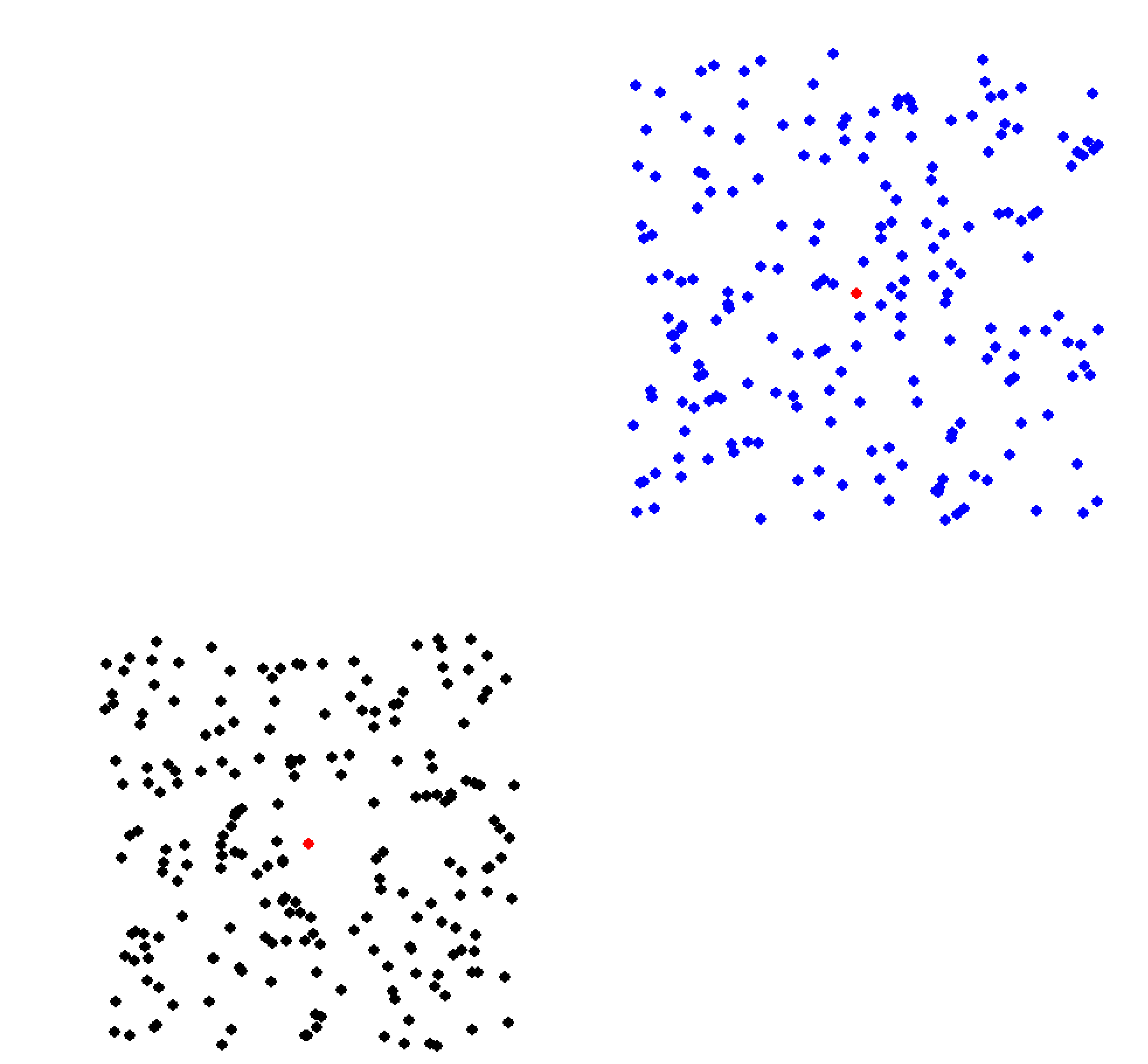
1. Выборка



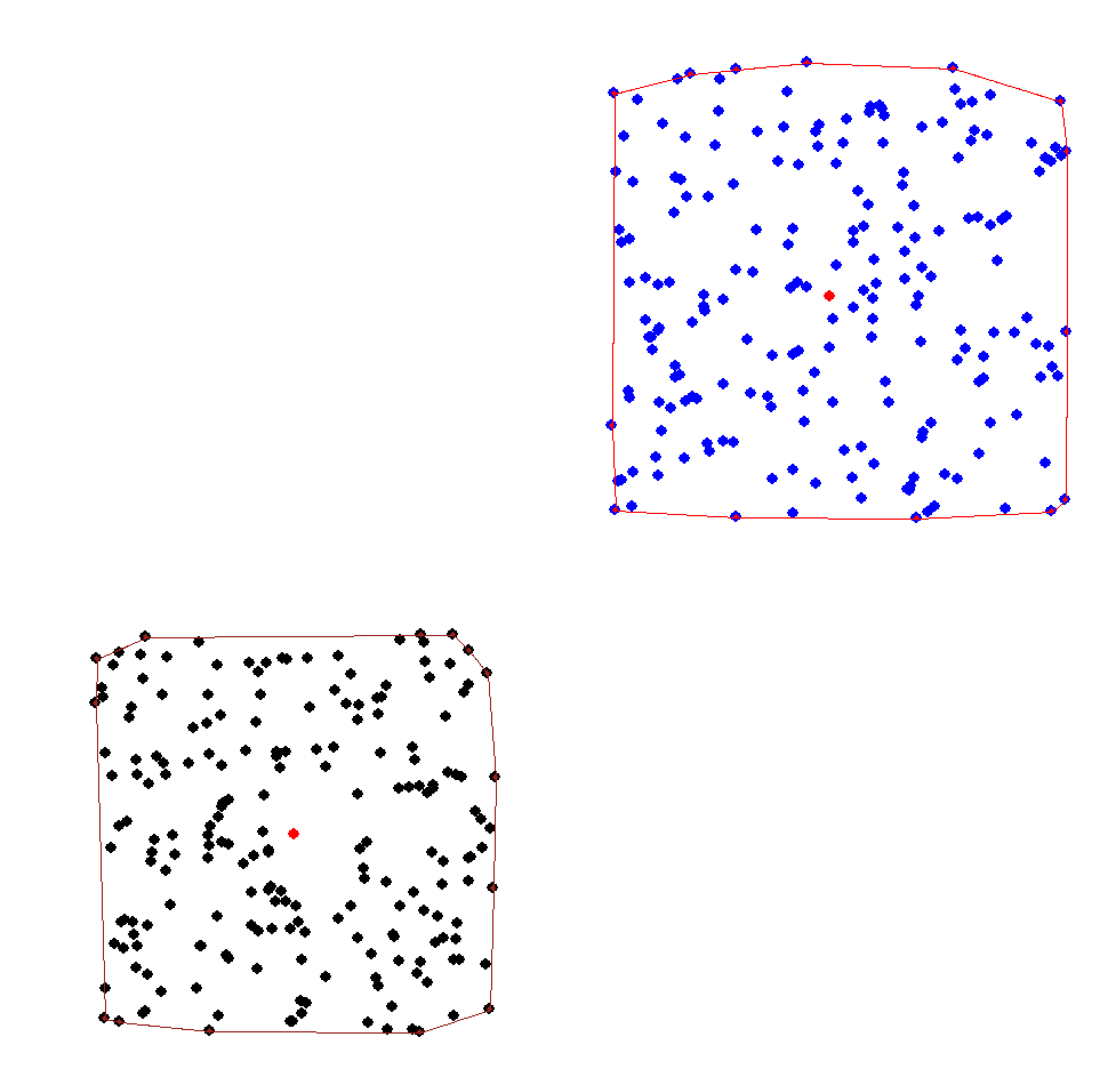
1. Центры кластеров на выборке



1. Полученные кластеры и их центры

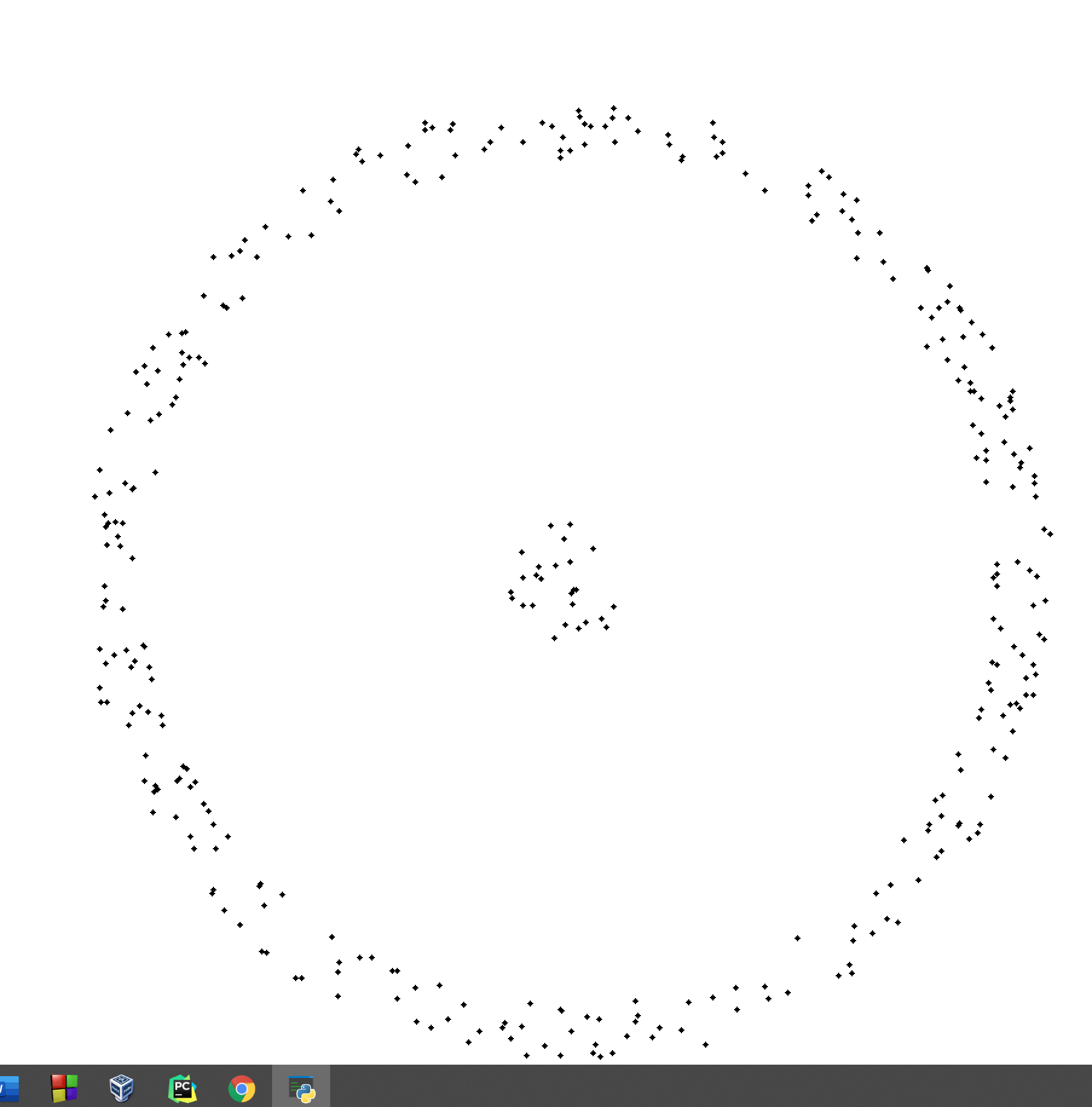


1. Выпуклая оболочка

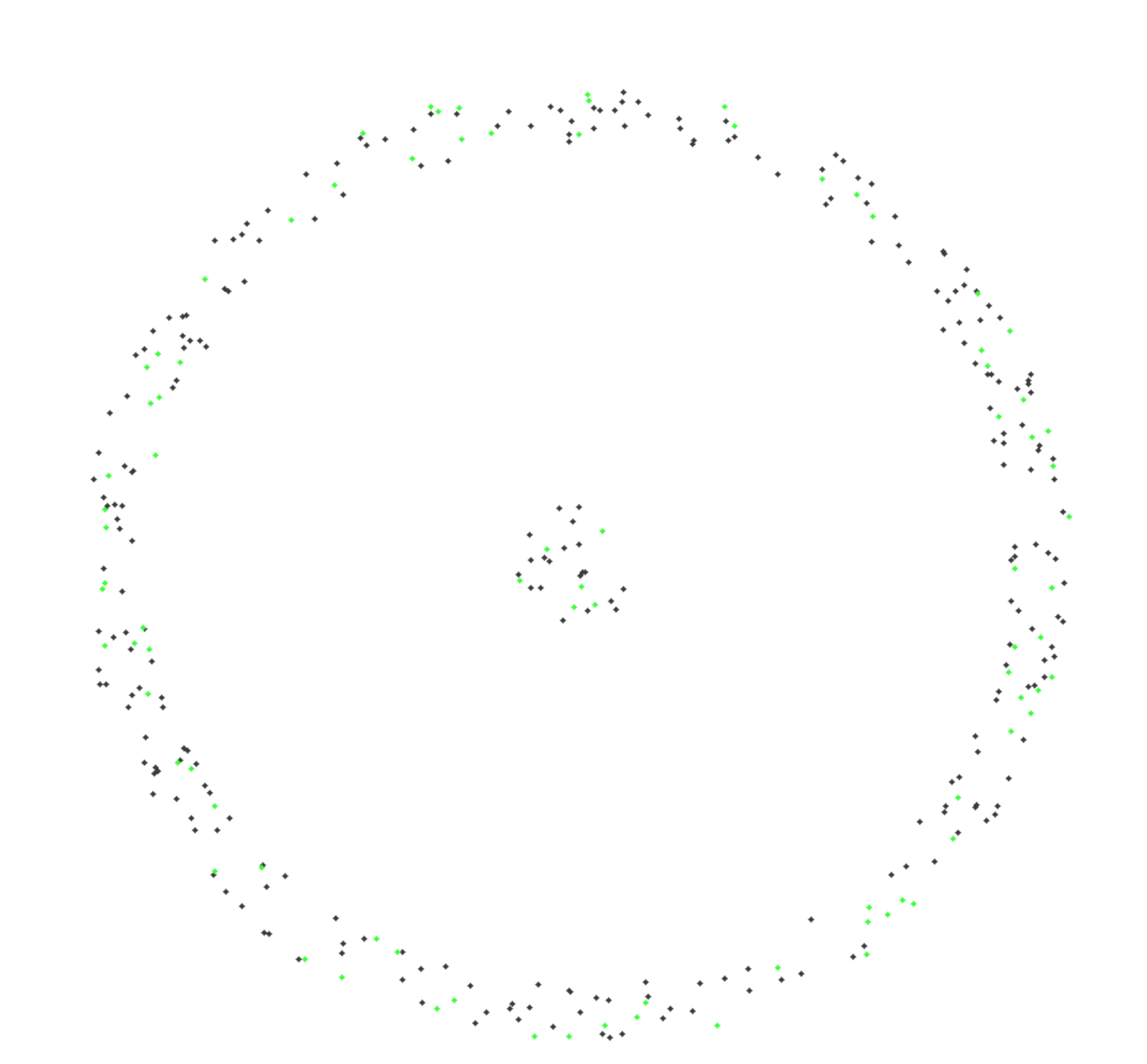


(вложенные кластеры)

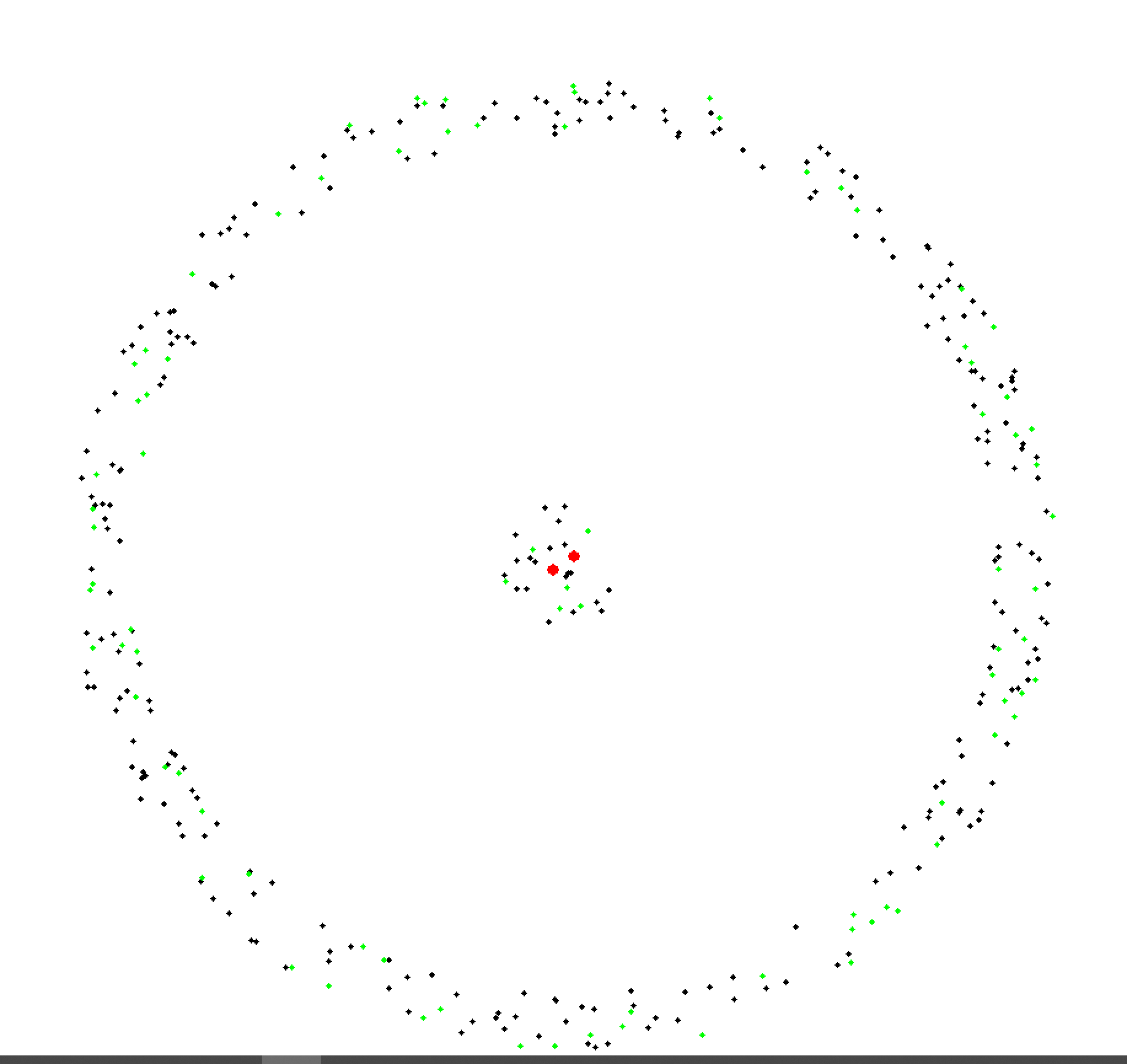
1. Генерация данных



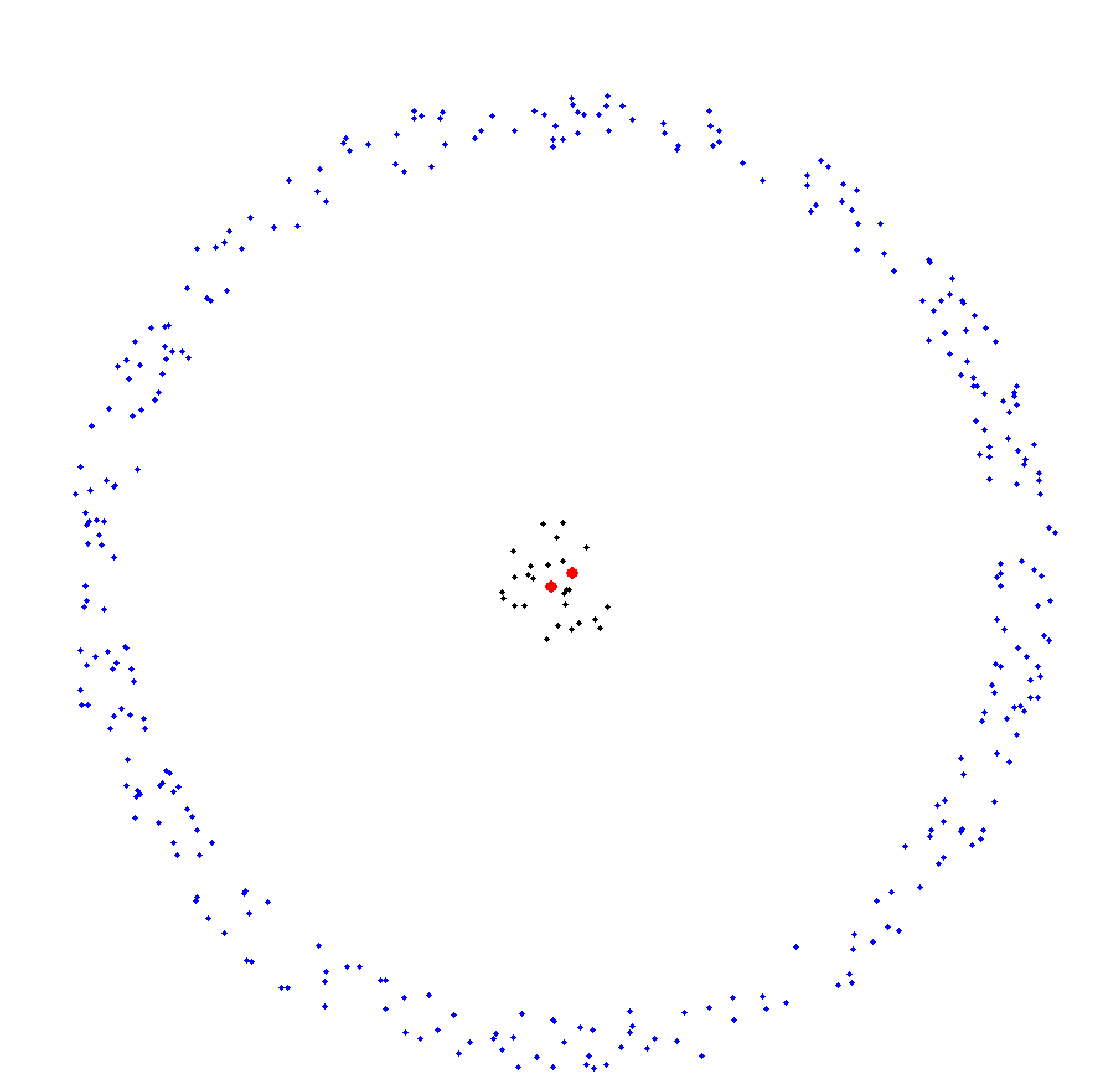
1. Выборка



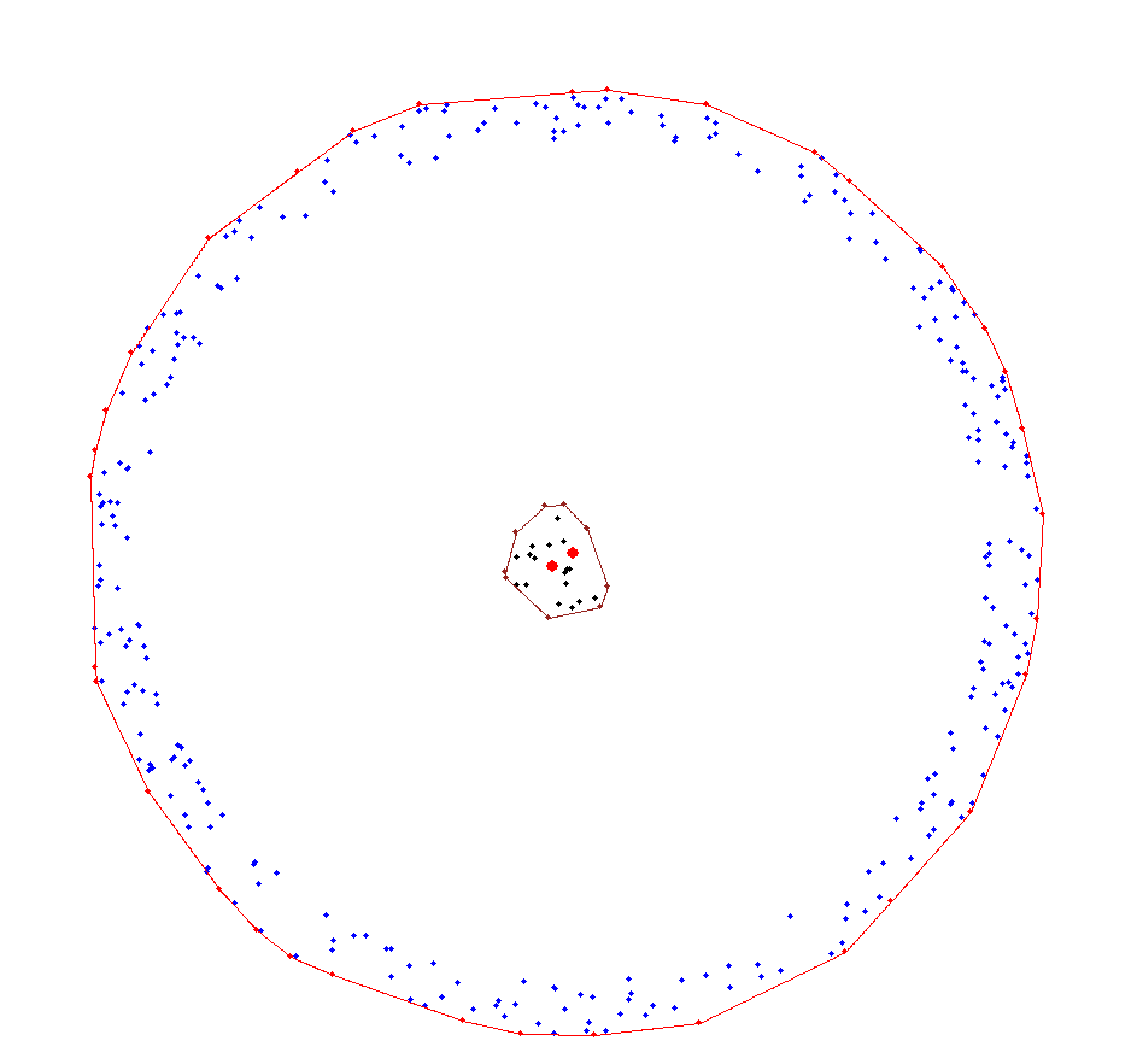
1. Центры кластеров на выборке



1. Полученные кластеры и их центры

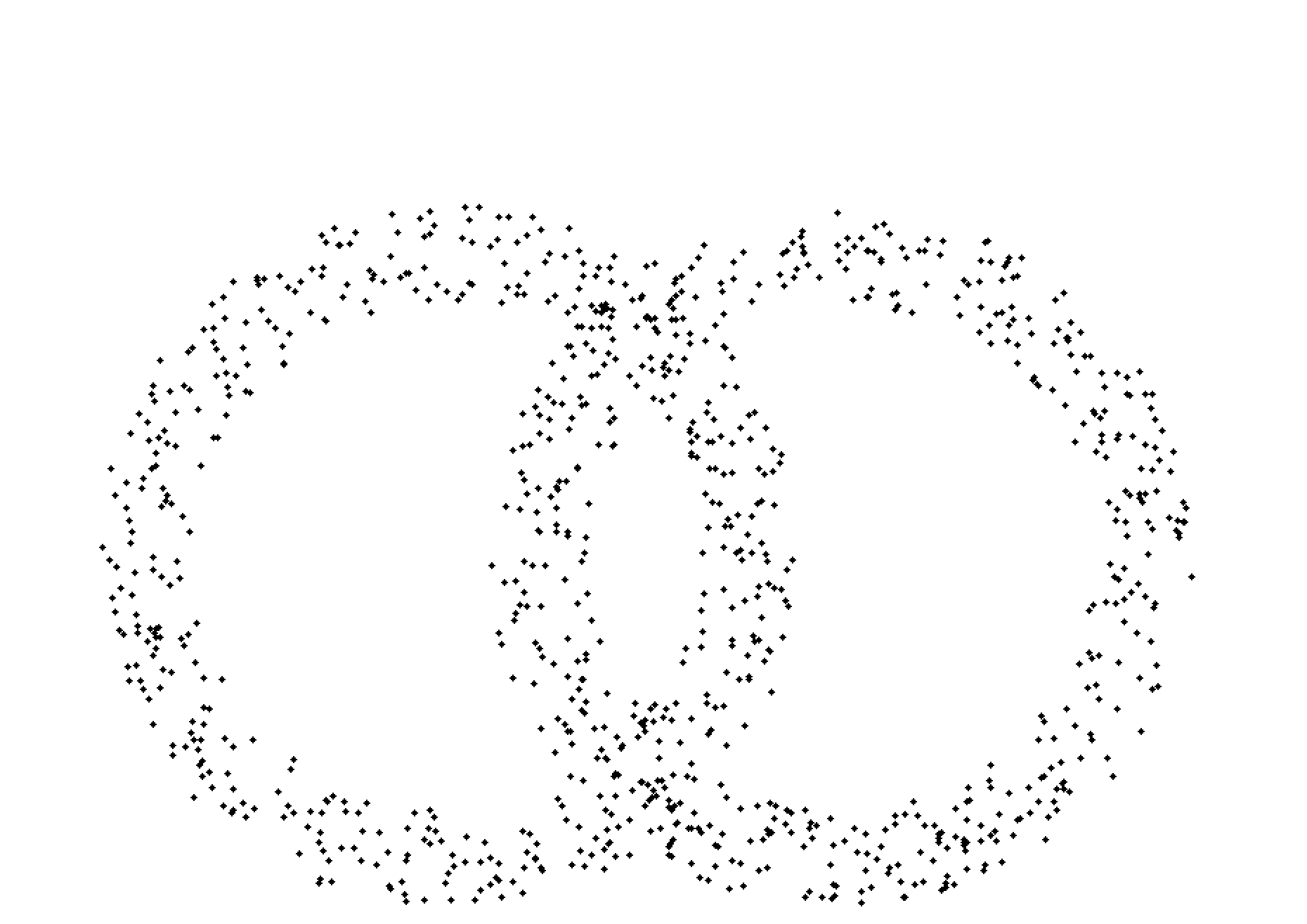


1. Выпуклая оболочка

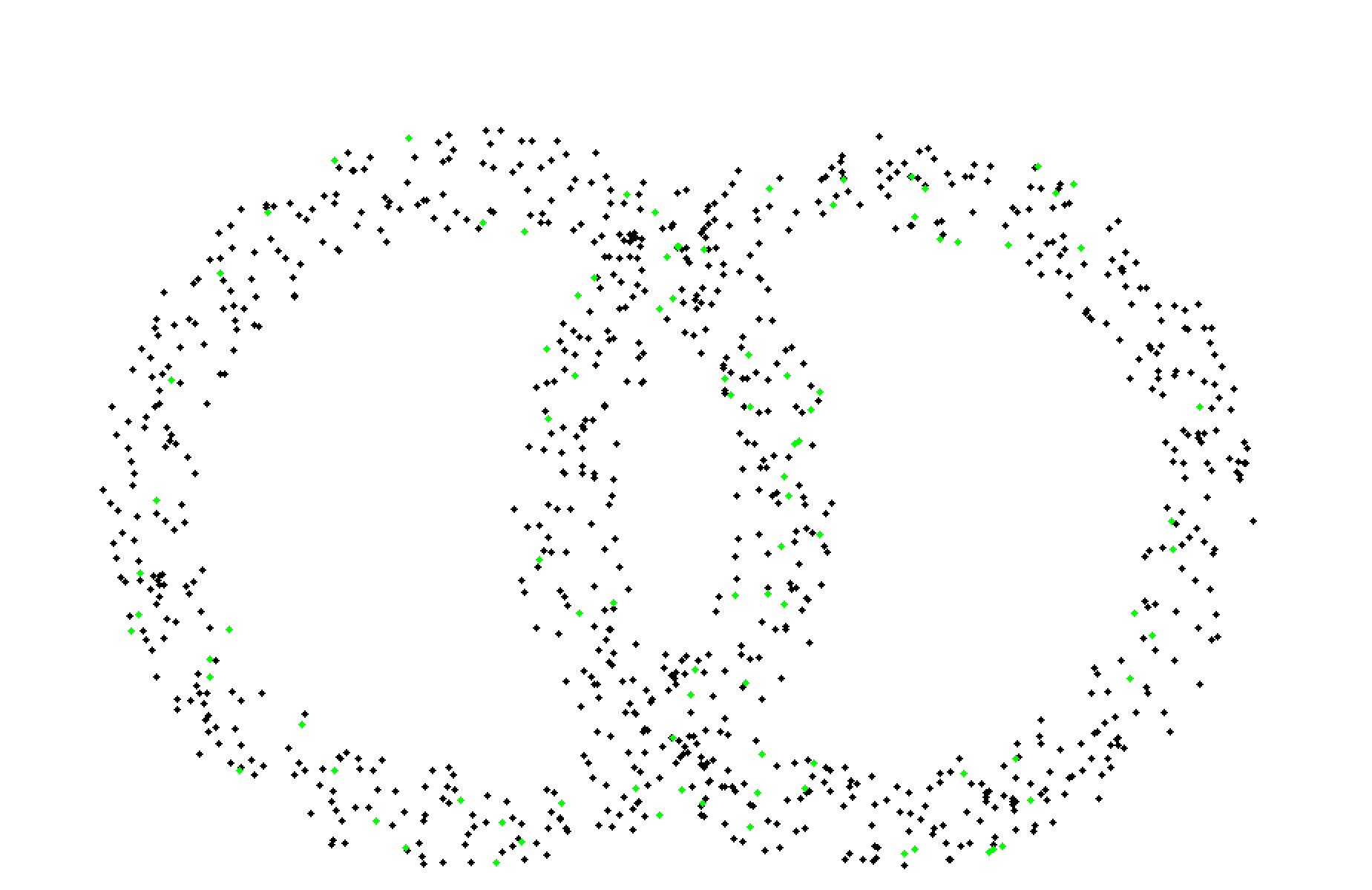


(пересечение двух колец)

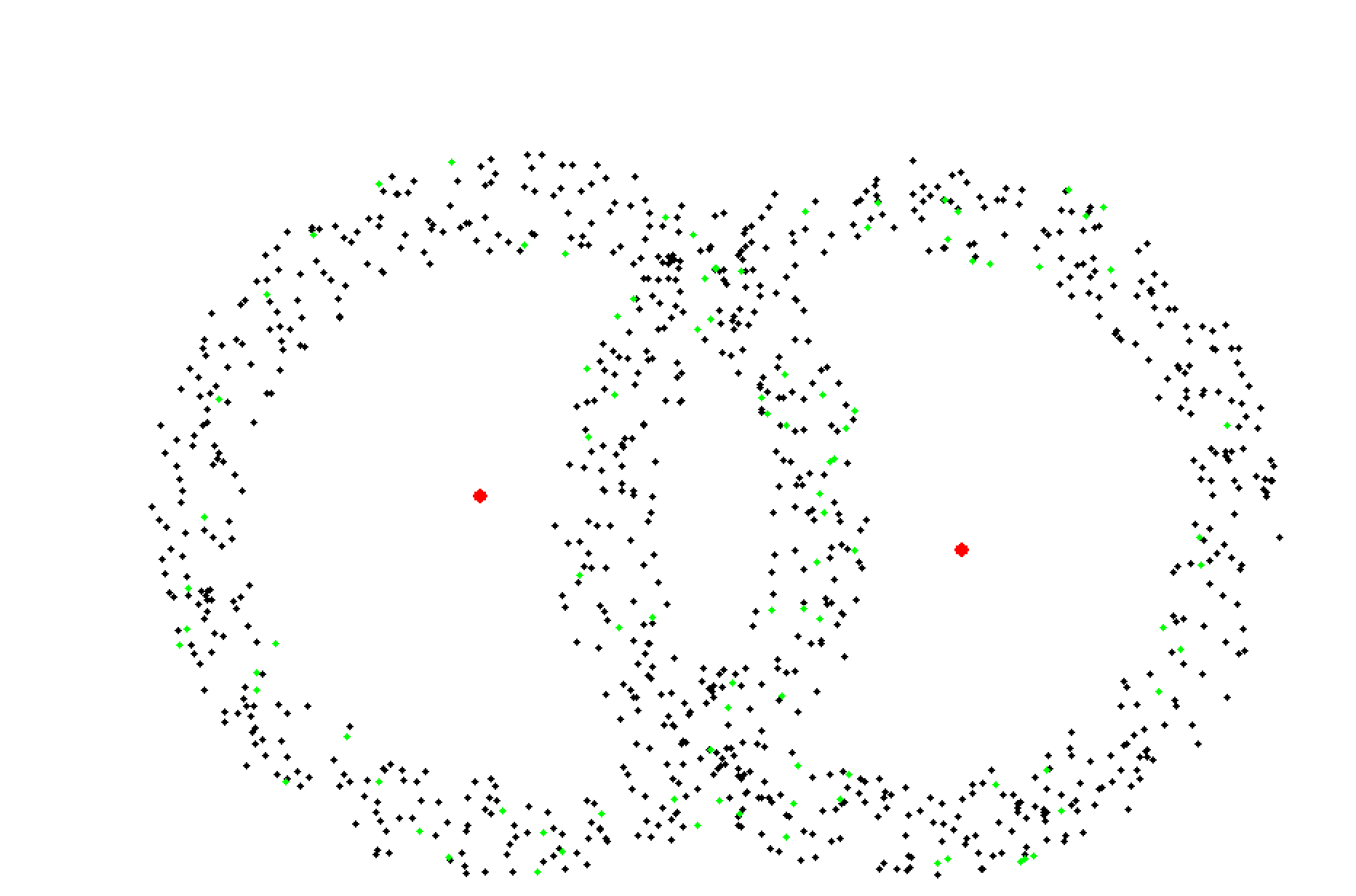
1. Генерация данных



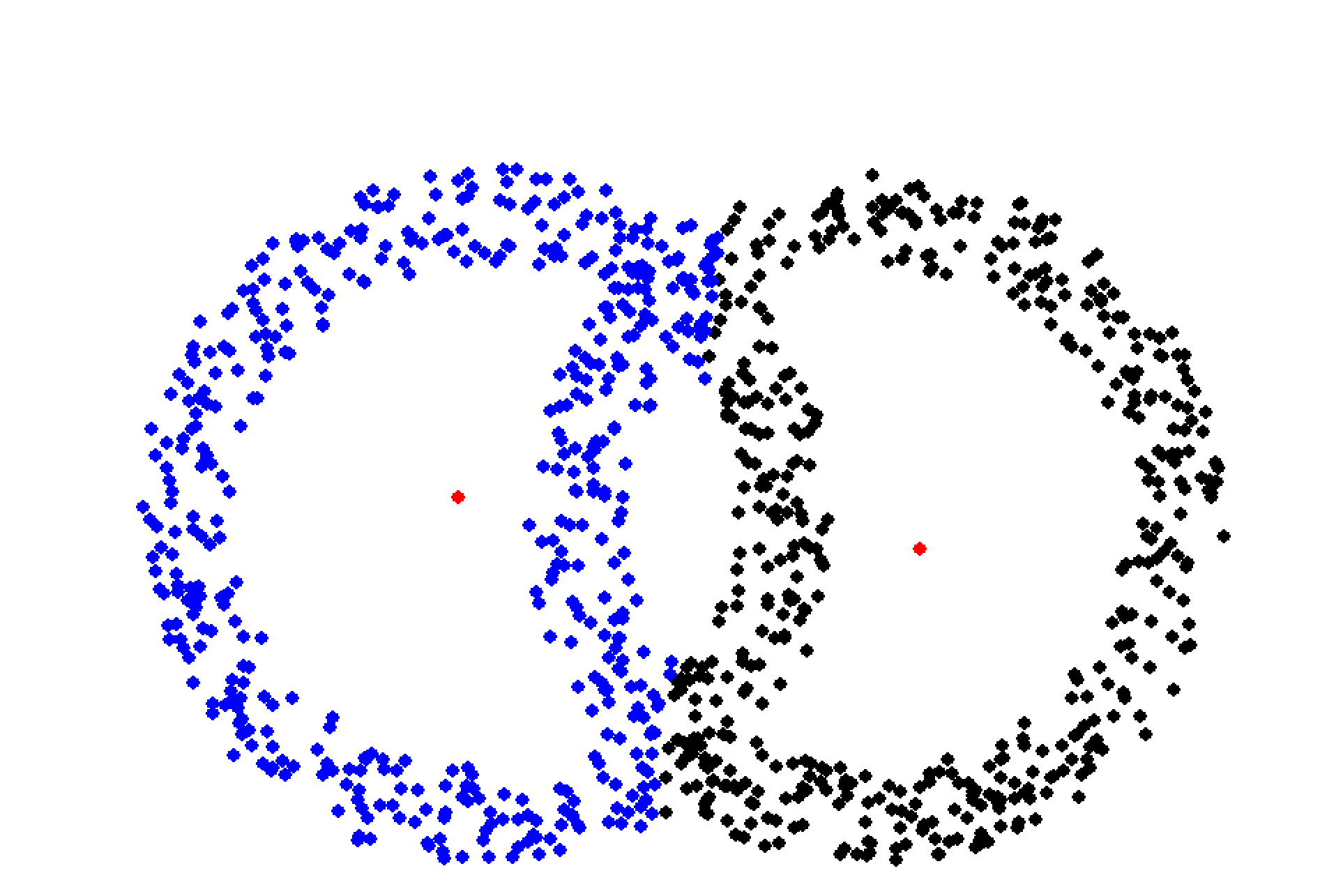
1. Выборка



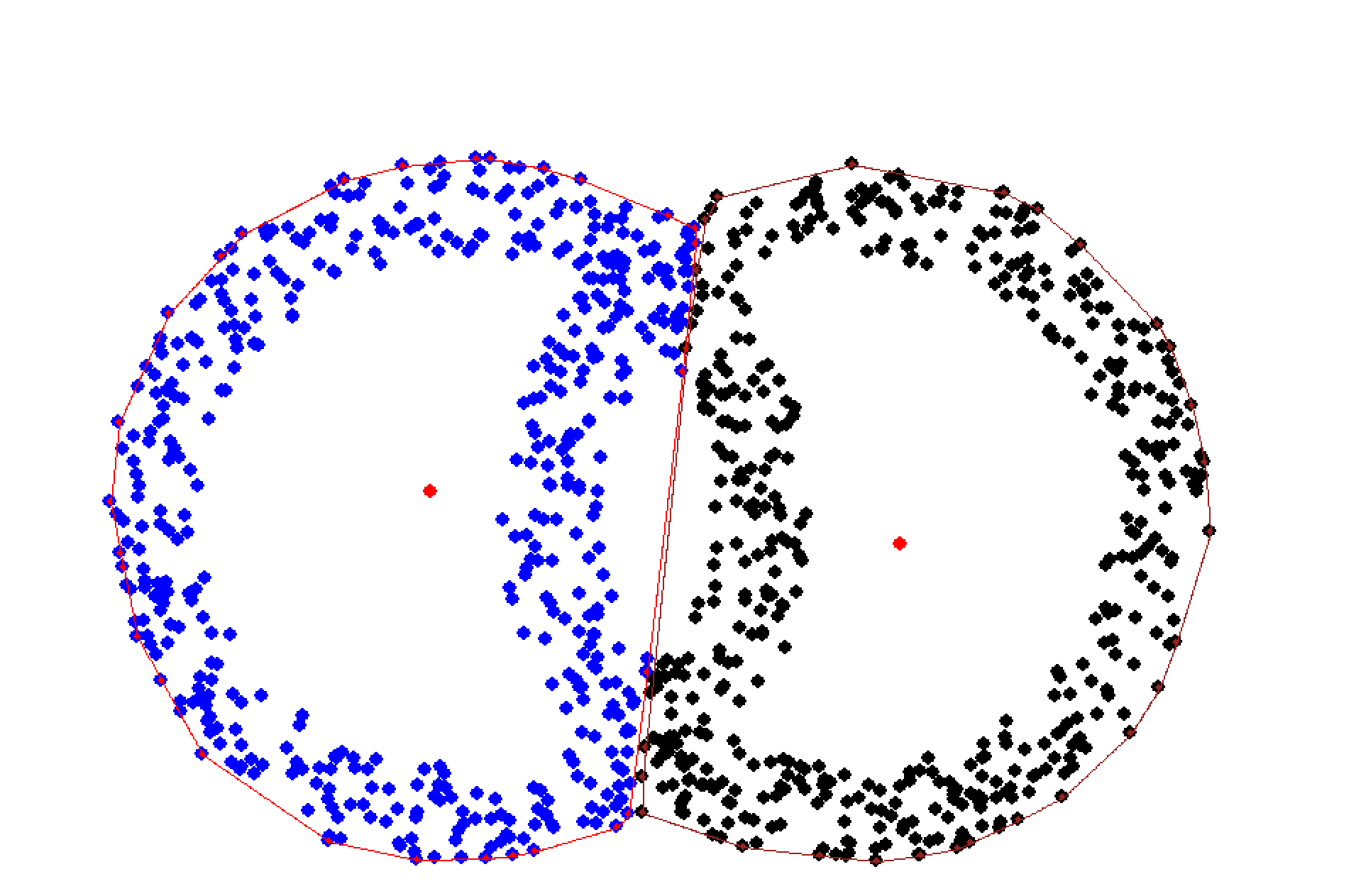
1. Центры кластеров на выборке



1. Полученные кластеры и их центры



1. Выпуклая оболочка



1. Выборка для k-средних

