

1. Кластеризация

Программа генерирует случайным образом 10000 точек с положительными координатами, не превосходящими 200. После чего из них составляется три датасета. Первый - данный набор точек в том виде, в котором он получился. Второй - шар в круге. Третий - два шара. После чего через k-means происходит разбиение всех точек на два кластера. Визуально это можно видеть на картинках.

В первом и третьем случаях разделение k-means выглядит адекватно. Что же о втором датасете, то там распределение на кластеры идет не совсем стандартным образом. Таким образом во втором случае использование k-means бессмысленно.