

Задание 8. Разделительная кластеризация

1. Выполните кластеризацию набора 2-х или 3-мерных данных с помощью алгоритма k-Means (предполагается, что полученные кластеры будут выпуклыми), используя различные значения параметра k (из интервала 3..9).

Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

2. Внесите шум в набор данных (случайным образом изменить определенную долю объектов набора: 1%, 3%, 5%, 10%; изменение может заключаться в добавлении/вычитании k /из одной/нескольких координат объекта случайного числа).

Выполните кластеризацию зашумленного набора данных с помощью алгоритмов k-Means и k-Medoids (или PAM), используя различные значения параметра k (из интервала 3..9).

Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

3. Выполните кластеризацию набора данных из задания 9 (с невыпуклыми кластерами) с помощью алгоритмов k-Means и k-Medoids (или PAM), используя различные значения параметра k (из интервала 3..9).

Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

Типовые контрольные вопросы

1. Объясните идею разделительной кластеризации и работу алгоритма k-Means.
2. Назовите преимущества и недостатки алгоритма k-Means.
3. Объясните работу алгоритма k-Medoids.