

Практическая работа №6

Статистический анализ

Группа 9372, 9373

Тема работы

Кластерный анализ. Метод k-средних

Цель работы

Освоение основных понятий и некоторых методов кластерного анализа, в частности, метода k-means.

Постановка задачи

Дано конечное множество из объектов, представленных двумя признаками (в качестве этого множества принимаем исходную двумерную выборку, сформированную ранее в практической работе №4). Выполнить разбиение исходного множества объектов на конечное число подмножеств (кластеров) с использованием метода k-means. Полученные результаты содержательно проинтерпретировать.

Порядок выполнения работы

1. Нормализовать множество точек из предыдущего раздела, отобразить полученное множество.
2. Определить «грубую» верхнюю оценку количества кластеров.
3. Реализовать алгоритм k-means. Отобразить полученные кластеры, выделить каждый кластер разным цветом, отметить центроиды.

4. Реализовать алгоритмы k-medians (для нечётных вариантов) и k-medoids (для чётных вариантов). Отобразить полученные кластеры, выделить каждый кластер разным цветом, отметить центроиды.
5. С помощью метода локтя и/или метода силуэтов выявить для каждого метода оптимальное количество кластеров.
6. На основании оценки качества разбиения определить наилучший метод кластеризации.
7. Сравнить методы. Сделать выводы.
- 8*. *Реализовать модификацию k-means++. Объяснить её преимущества. Сравнить с обычным методом k-means.*

Вопросы для самоконтроля

1. Сформулировать основные задачи кластерного анализа.
2. Дать классификацию и охарактеризовать основные методы кластерного анализа.
3. Критерии качества кластерных разбиений.
4. Способы выбора количества кластеров для выборки.
5. Описать и прокомментировать метод k-means кластерного анализа.
6. Описать и прокомментировать метод k-medians кластерного анализа.
7. Описать и прокомментировать метод k-medoids кластерного анализа.