Лабораторная работа №6: Кластерный анализ. Метод k-средних

Цель работы

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

Постановка задачи

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

Порядок выполнения работы

1. Нормализовать множество точек, отобразить полученное множество.
2. Определить верхнюю оценку количества кластеров по формуле: $ \bar k = \lfloor\sqrt{N/2}\rfloor $, где $ N $ – число точек.
3. Реализовать алгоритм k-means в двух вариантах: пересчет центра кластера осуществляется после каждого изменения его состава и пересчет центра кластера осуществляется лишь после того, как будет завершен просмотр всех данных (шаг процедуры). Отобразить полученные кластеры, выделить каждый кластер разным цветом, отметить центроиды.
4. Провести оценку качества разбиения для различных разбиений.
5. Содержательно проинтерпретировать полученные результаты.
6. *Дополнительные необязательные задания:*
    1. *Реализовать алгоритмы k-medians и k-medoids. Отобразить полученные кластеры, выделить каждый кластер разным цветом, отметить центроиды. Провести оценку методов, сделать выводы.*
    2. *С помощью метода локтя и/или метода силуэтов выявить для каждого метода оптимальное количество кластеров.*
    3. *Реализовать модификацию k-means++. Объяснить её приемущества. Сравнить с обычным методом k-means.*

Содержание отчёта

1. Цель работы.
2. Краткое изложение основных теоретических понятий.
3. Постановка задачи с кратким описанием порядка выполнения работы.

4. Необходимые формулы, рисунки и таблицы.
5. Краткие выводы по полученным результатам.
6. Общий вывод по проделанной работе.
7. Код программы (если имеется).

Вопросы для самоконтроля

1. Сформулировать основные задачи кластерного анализа.
2. Дать классификацию и охарактеризовать основные методы кластерного анализа.
3. Критерии качества кластерных разбиений.
4. Описать и прокомментировать метод k-средних кластерного анализа.

From:
http://se.moevm.info/ - **se.moevm.info**

Permanent link:
http://se.moevm.info/doku.php/courses:statistical_methods_of_experimental_data_handling:prac6?rev=1617706385

Last update: **2022/12/10 09:08**