Практическая работа №6: 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: $k = \frac{N}2$ γгисло точек.
- 3. Реализовать алгоритм k-means в двух вариантах:
 - 1. пересчет центра кластера осуществляется после каждого изменения его состава;
 - 2. пересчет центра кластера осуществляется лишь после того, как будет завершен просмотр всех данных (шаг процедуры).
- 4. На каждом шаге процедуры разбиения методом k-means вычислять функционалы качества полученного разбиения:
 - 1. \$ F_1 \$ сумма по всем кластерам квадратов расстояний элементов кластеров до центров соответствующих кластеров;
 - 2. \$ F_2 \$ сумма по всем кластерам внутрикластерных расстояний между элементами кластеров;
 - 3. \$ F_3 \$ сумма по всем кластерам внутрикластерных дисперсий (относительно центров кластеров).
- 5. Отобразить полученные кластеры, выделить каждый кластер разным цветом, отметить центроиды.
- 6. Содержательно проинтерпретировать полученные результаты.
- 7. 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-means кластерного анализа.

From:

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

Permanent link:

 $http://se.moevm.info/doku.php/courses:statistical_methods_of_experimental_data_handling:prac6?rev=165644022221.$

Last update: 2022/12/10 09:08

http://se.moevm.info/ Printed on 2024/06/01 06:00