

Исследование алгоритмов кластеризации

Цель работы

Исследовать методов кластеризации на примере алгоритмов иерархической группировки и k-средних (k-means).

Форма контроля

Письменный отчёт (допускается представление в электронном виде). Опрос в устной форме в соответствии с перечнем контрольных вопросов.

Количество отведённых аудиторных часов

4

Содержание работы

Получить у преподавателя вариант задания и написать код, реализующий соответствующий алгоритм обработки информации. Для ответа на поставленные в задании вопросы провести численный эксперимент или статистическое имитационное моделирование и представить соответствующие графики. Провести анализ полученных результатов и представить его в виде выводов по проделанной работе.

Пример варианта задания

1. Реализовать классификацию объектов 3х классов на основе иерархического алгоритма. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
2. Реализовать классификацию объектов 4х классов на основе иерархического алгоритма. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
3. Реализовать классификацию объектов 5ти классов на основе иерархического алгоритма. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
4. Реализовать классификацию объектов 3х классов на основе алгоритма k-средних. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
5. Реализовать классификацию объектов 4х классов на основе алгоритма k-средних. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
6. Реализовать классификацию объектов 5ти классов на основе алгоритма k-средних. Выбрать метрику (функцию расстояния), минимизирующую ошибку классификации.
7. Сравнить иерархический алгоритм и k-means по критерию вероятности ошибки классификации при использовании определённых метрик: абсолютной, Евклидовой.
8. Сравнить иерархический алгоритм и k-means по критерию вероятности ошибки классификации при использовании определённых метрик: Евклидовой, Махаланобиса.

Примеры контрольных вопросов

1. Что такое дендрограмма?
2. Какой критерий минимизируется в алгоритме k-средних?