Исмакова Даяна

Летучка кластеры

1. Что такое кластерный анализ? В каких случаях он применяется?

Кластерный анализ – это своего рода классификация без обучения, т.е. изначально мы не знаем какие будут классы или кластеры, получается, что мы разделяем заданную выборку объектов на подмножества или кластеры. Используется для группирования ряда объектов с последующими проверками.

1. Охарактеризуйте входные данные кластерного анализа.

Входные данные кластерного анализа должны соответствовать следующим требованиям:

* 1. показатели не должны коррелировать между собой;
  2. показатели не должны противоречить теории измерений;
  3. распределение показателей должно быть близко к нормальному;
  4. показатели должны отвечать требованию «устойчивости», под которой понимается отсутствие влияния на их значения случайных факторов;
  5. выборка должна быть однородна, не содержать «выбросов».

1. Приведите пример (формулу) вычисления расстояния между группами объектов.

Приведу одну из формул для вычисления расстояния, а конкретно расстояние ближайшего соседа:

1. Из каких шагов состоит алгоритм кластеризации?  
   В основном алгоритмы кластерного анализа состоят из 2-ух этапов:
   1. задается начальное (возможно, искусственное или даже произвольное) разбиение множества объектов на классы и определяется некоторый математический критерий качества автоматической классификации.
   2. объекты переносятся из класса в класс до тех пор, пока значение критерия не перестанет улучшаться.
2. Перечислите и кратко охарактеризуйте методы кластеризации.
   1. Вероятностный подход: предполагается, что каждый рассматриваемый объект относится к одному из к-классов. (нап., можно использовать K-средних (K-means))
   2. Теоретико-графовый подход. При таком подходе исходят из того, что некое множество объектов характеризуется определённой степенью связности.
   3. Логический подход. В таком случае происходит построение дендрограммы с помощью дерева решений. (Под дендрограмой обычно понимается дерево, т.е. граф без циклов построенный по матрице мер близости.)