

Особенности методов кластерного анализа

Введение в кластерный анализ

Кластерный анализ

Кластерный анализ — методы разбиения заданной выборки объектов (ситуаций) на непересекающиеся подмножества (кластеры), так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

Объекты группы должны обладать схожими характеристиками.

Мера сходства — показатель сходства характеристик сравниваемых объектов.

Кластерный анализ

Выборки

$$Y_n = \{Y_1, \dots, Y_n\}$$

$$Y_i \leftarrow (X_1^i, \dots, X_m^i)$$

Кластерный анализ

! Нельзя использовать абсолютно все данные для выделения групп.

Необходимо выделять только те признаки, которые помогут эффективно разделить объекты на группы.

Особенности методов кластерного анализа

1. Эмпирические методы без статистической базы.
2. Методы кластерного анализа используют в самых различных научных дисциплинах. Общий вопрос исследователей: как организовать наблюдаемые данные в наглядные структуры?
3. Разные методы будут давать различные результаты для одних и тех же данных.
4. Кластерный анализ может искажать результаты структурирования и выделять несуществующие группы.

Меры сходства

- Меры расстояния
- Корреляционные меры сходства
- Меры сходства ассоциативности (для бинарных данных)
- Вероятностные меры сходства

Иерархический кластерный анализ

Иерархическая кластеризация — последовательное объединение меньших кластеров в большие или разделение больших кластеров на меньшие.

Итерационные методы кластерного анализа

1. Метод k-средних
2. FOREL