

Способы оценки качества кластеризации

Итерационные методы кластерного анализа

Способы оценки качества кластеризации

1. Статистические

Формальные:

- Скорректированный R-квадрат.
- Дисперсионный анализ.

«Эмпирические»:

- Наполненность кластеров.
- Устойчивость решения.

2. Содержательные

- Интерпретируемость построенной модели.
- Согласованность с экспертными оценками.

Формальные методы оценки качества кластеризации

- Скорректированный R-квадрат.

$$R^2_{adj} = 1 - \frac{WSS \cdot (n-1)}{TWSS \cdot (n-k)}$$

- Дисперсионный анализ.
- Критерий, основанный на соотношении внутрикластерных и межкластерных различий.
- Коэффициент разбиения и основанный на нём индекс чёткости.
- Показатель компактности и изолированности.