

Статистический анализ данных. Спецкурс.

Лекция 5. Классификация

Ботанический сад-институт ДВО РАН

Кислов Д.Е.
27 ноября 2016 г.

- Классификация в отсутствии обучающей выборки (кластеризация);

Задачи классификации

- Классификация в отсутствии обучающей выборки (кластеризация);
- Классификация при наличии обучающей выборки (по прецедентам);

- Иерархический (агломеративная, дивизивные);

- Иерархический (агломеративная, дивизивные);
- Логическая кластеризация

- Иерархический (агломеративная, дивизивные);
- Логическая кластеризация
- Вероятностные

- Иерархический (агломеративная, дивизивные);
- Логическая кластеризация
- Вероятностные
- Нейросетевые

Общая структура алгоритмов

- задание расстояние между кластеризуемыми объектами;
- задание расстояние между группами объектов;

Полагается, что объекты x, y имеют координаты x_1, \dots, x_n и y_1, \dots, y_n .

- Евклидово расстояние: $\rho(x, y) = \sum_j (x_j - y_j)^2$;

Расстояние между объектами

Полагается, что объекты x, y имеют координаты x_1, \dots, x_n и y_1, \dots, y_n .

- Евклидово расстояние: $\rho(x, y) = \sum_j (x_j - y_j)^2$;
- Расстояние Чебышева: $\rho(x, y) = \max_j |x_j - y_j|$;

Расстояние между объектами

Полагается, что объекты x, y имеют координаты x_1, \dots, x_n и y_1, \dots, y_n .

- Евклидово расстояние: $\rho(x, y) = \sum_j (x_j - y_j)^2$;
- Расстояние Чебышева: $\rho(x, y) = \max_j |x_j - y_j|$;
- Расстояние city-block: $\rho(x, y) = \sum_i |x_i - y_i|$;

Расстояние между объектами

Полагается, что объекты x, y имеют координаты x_1, \dots, x_n и y_1, \dots, y_n .

- Евклидово расстояние: $\rho(x, y) = \sum_j (x_j - y_j)^2$;
- Расстояние Чебышева: $\rho(x, y) = \max_j |x_j - y_j|$;
- Расстояние city-block: $\rho(x, y) = \sum_i |x_i - y_i|$;
- Расстояние Минковского: $\rho(x, y)^p = \sum_i (x_i - y_i)^p$;

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);
- попарное среднее;

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);
- попарное среднее;
- центроидный метод;

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);
- попарное среднее;

Подходы для вычисления расстояний между группами объектов

- метод минимального расстояния (single method);
- метод максимального расстояния (complete method);
- попарное среднее;
- центроидный метод;

Представление иерархической кластеризации в виде дендрограммы

