

# Задача и алгоритмы кластеризации

Николай Анохин

# Задача кластеризации

В задачах кластеризации значение целевой функции для объектов из обучающей выборки неизвестно. Цель кластеризации – исследовать “скрытую структуру” данных.

## Постановка задачи

**Дано.** Признаковые описания  $N$  объектов  $\mathbf{x} \in \mathcal{X}$ , образующие тренировочный набор данных  $X$

**Найти.** Модель из семейства параметрических функций

$$H = \{h(\mathbf{x}, \theta) : \mathcal{X} \times \Theta \rightarrow \mathcal{Y} \mid \mathcal{Y} = \{1, \dots, K\}\},$$

ставящую в соответствие произвольному  $\mathbf{x} \in \mathcal{X}$  один из  $K$  кластеров так, чтобы объекты внутри одного кластера были похожи, а объекты из разных кластеров различались