

## Лекция 1. Введение.

Объект ( $X$ ) — то, что анализируем.

Целевая переменная ( $Y$ ) — то, что предсказываем.

Обучающая выборка  $X = \{(x_i, y_i)\}_{i=1}^{\ell}$  — набор объектов, для которого мы знаем правильные ответы.

Признаки, факторы — характеристики объектов.

## Признаки:

★ числовые,  $x_j \in \mathbb{R}$

★ категориальные,  $x_j \in \{c_1, \dots, c_m\}$

★ порядковые,  $x_j \in \{c_1, \dots, c_m\}$ , но есть операция сравнения  $<$

Типы задач (определяется целевой переменной  $Y$ ):

★ Обучение с учителем (*supervised learning*) — есть целевая переменная.

1.  $Y = \mathbb{R}$  — регрессия.

2.  $|Y| < \infty$  — классификация.

( $Y = \{0, 1\}$  — бинарная,  $Y = \{1, \dots, k\}$  — многоклассовая классификация,

$Y = \{0, 1\}^k$  — классификация с пересекающимися классами)

★ Обучение без учителя (*unsupervised learning*) — целевой переменной нет.

1. Кластеризация:  $X$  надо разбить на группы, чтобы в каждой группе объекты были похожи.

2. Оценивание плотности.

★ Частичное обучение.

★ Обучение с подкреплением.

Модель, алгоритм:  $a: X \rightarrow Y$ .

Примеры:  $a(x) = w_0 + w_1x + \dots + w_dx_d$

$\mathcal{A} = \{a(x, w) | w\}$  — семейство моделей.

Функция потерь (*loss function*):  $L: Y \times Y \rightarrow \mathbb{R}$ .

Функционал ошибки:  $Q(a, X)$ . В большинстве случаев:  $Q(a, X) = \frac{1}{\ell} \sum_{i=1}^{\ell} L(y_i, a(x_i))$

Обучение модели:  $Q(a, X) \rightarrow \min_{a \in \mathcal{A}}$