

**Методы анализа данных**  
**Александр Широков ПМ-1701**

Преподаватель:

ИВАХНЕНКО ДАРЬЯ АЛЕКСАНДРОВНА

Санкт-Петербург  
2020 г., 7 семестр

## Список литературы

[1]

## Содержание

<b>1</b>	<b>01.09.2020</b>	<b>2</b>
1.1	Задача обучения по прецедентам . . . . .	2
1.2	Типы задач . . . . .	2

# 1 01.09.2020

## 1.1 Задача обучения по прецедентам

Пусть  $X$  - множество объектов, а  $Y$  - множество ответов.  $y : X \rightarrow Y$  - неизвестная зависимость.

Дано:  $\{x_1, \dots, x_l\} \subset X$  - обучающая выборка, а  $y_i = y(x_i), i = 1, \dots, l$  - известные ответы.

Требуется найти  $a : X \rightarrow Y$  - алгоритм, решающую функцию, приближающую  $y$  на всем множестве  $X$ .

## 1.2 Типы задач

### Задачи восстановления регрессии:

- $Y = \mathbb{R}$  - вся числовая ось:
  - определение температуры воздуха метеорологического поля
  - оценка влияния факторов потребления
- $Y \in [0; +\infty)$ :
  - задачи медицинской диагностики: прогнозирование ожидаемого время действия препарата
  - задачи кредитного скоринга: определение величины кредитного лимита
  - определение расхода топлива по техническим характеристикам
- $Y \in [0, 1, \dots, +\infty)$  - счетная целевая переменная

### Задача классификации:

- $Y = \{-1, +1\}$  - классификация на два класса:
  - задачи кредитного скоринга: решение о выдаче кредита
  - предсказание оттока клиентов
- $Y = \{1, \dots, K\}$  - классификация на  $K$  непересекающихся классов:
  - задачи медицинской диагностики: определение диагноза

- распознавание символов
- определение жанра
- $Y = \{0, 1\}^K$  - на  $K$  классов, которые могут пересекаться:
  - определение ключевых слов для оптимизации поиска
  - определение присутствующих на фото объектов

### **Типы признаков**

- $D_j = \{0, 1\}$  - бинарный признак  $f_j$ :
  - пол
  - является ли..?
- $|D_j| < \infty$  - номинальный признак  $f_j$ :
  - город
  - цвет
- $|D_j| < \infty$ ,  $D_j$  - упорядочено - порядковый признак  $f_j$ :
  - уровень холестерина (ниже нормы, норма, выше нормы)
- $D_j = \mathbf{R}$  - количественный признак  $f_j$ :
  - длина и ширина объекта