

# Лекция 1. Введение в машинное обучение

Александр Юрьевич Авдюшенко

МКН СПбГУ, ШАД

14 июня 2022



**Факультет  
математики  
и компьютерных  
наук  
СПбГУ**

# Первый data scientist

математик, механик, физик, астроном и геодезист

# Первый data scientist

математик, механик, физик, астроном и геодезист

В 24 года предсказал, где искать малую планету *Цереру*,  
скрывшуюся за Солнцем

# Первый data scientist

математик, механик, физик, астроном и геодезист

В 24 года предсказал, где искать малую планету *Цереру*, скрывшуюся за Солнцем



## Определение из Википедии

**Машинное обучение** (англ. machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач.

Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

Что это?

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u_i)}{\partial x_i} = 0$$
$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial[\rho u_i u_j]}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho f_i$$

# Что это?

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u_i)}{\partial x_i} = 0$$
$$\frac{\partial(\rho u_i)}{\partial t} + \frac{\partial[\rho u_i u_j]}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho f_i$$

В машинном обучении нет предзаданной модели с уравнениями...

# Два типа обучения

- ▶ Обучение по прецедентам (обучение с учителем), или индуктивное обучение, основано на выявлении эмпирических закономерностей в данных.
- ▶ Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Дедуктивное обучение принято относить к области *экспертных систем*, поэтому *машинное обучение*  $\sim$  *обучение по прецедентам*.

Многие методы машинного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с извлечением информации (англ. information extraction, information retrieval), интеллектуальным анализом данных (data mining).



● data science  
Поисковый запрос

● machine learning  
Поисковый запрос

● data mining  
Поисковый запрос

+ Добавить сравнение

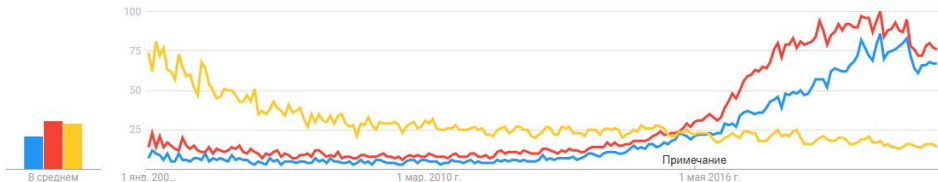
По всему миру ▼

2004 – настоящее время ▼

Все категории ▼

Веб-поиск ▼

Динамика популярности ?



# Математическая постановка

$X$  — множество объектов

$Y$  — множество ответов

$y : X \rightarrow Y$  — неизвестная зависимость (target function)

Задача по обучающей выборке (training sample)

$\{x_1, \dots, x_\ell\} \subset X$

с известными ответами  $y_i = y(x_i)$

найти

$a : X \rightarrow Y$  — алгоритм,

решающую функцию (decision function), приближающую  $y$  на всём множестве  $X$

Всё машинное обучение про это:

- ▶ как задаются объекты  $x_i$  и какими могут быть ответы  $y_i$
- ▶ в каком смысле « $a$  приближает  $y$ »
- ▶ как строить функцию  $a$

# Объекты и их признаки

$$f_j : X \rightarrow D_j$$

Вектор  $(f_1(x), \dots, f_n(x))$  — признаковое описание объекта  $x$

Типы признаков:

- ▶  $D_j = \{0, 1\}$  — бинарный
- ▶  $\#|D_j| < \infty$  — категориальный (номинальный)
- ▶  $\#|D_j| < \infty, D_j$  упорядочено — ординальный (порядковый)
- ▶  $D_j = \mathbb{R}$  — вещественный (количественный)

Матрица «объекты-признаки» (feature data)

$$F = ||f_j(x_i)||_{\ell \times n} = \begin{bmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_n(x_\ell) \end{bmatrix}$$

## Вопрос

*Как перевести все признаки в бинарные?*

# Типы задач

## Классификация (classification)

- ▶  $Y = \{-1, +1\}$  — бинарная классификация
- ▶  $Y = \{1, \dots, M\}$  — многоклассовая классификация
- ▶  $Y = \{0, 1\}^M$  — многоклассовая с пересекающимися классами

## Регрессия (regression)

$$Y = \mathbb{R} \text{ или } Y = \mathbb{R}^m$$

## Ранжирование (ranking)

$Y$  — конечное упорядоченное множество

# Предсказательная модель

Модель (predictive model) — параметрическое семейство функций

$$A = \{g(x, \theta) | \theta \in \Theta\},$$

где  $g : X \times \Theta \rightarrow Y$  — фиксированная функция,  $\Theta$  — множество допустимых значений параметра  $\theta$

## Пример

Линейная модель с вектором параметров

$$\theta = (\theta_1, \dots, \theta_n), \Theta = \mathbb{R}^n:$$

$$g(x, \theta) = \sum_{j=1}^n \theta_j f_j(x) \text{ — для регрессии и ранжирования, } Y = \mathbb{R}$$

$$g(x, \theta) = \text{sign} \sum_{j=1}^n \theta_j f_j(x) \text{ — для классификации, } Y = \{-1, +1\}$$

## Этап обучения (train)

Метод  $\mu$  по выборке  $(X, Y) = (x_i, y_i)_{i=1}^{\ell}$  строит алгоритм  $a = \mu(X, Y)$

$$\boxed{\begin{bmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_{\ell}) & \dots & f_n(x_{\ell}) \end{bmatrix}} \xrightarrow{y} \begin{bmatrix} y_1 \\ \dots \\ y_{\ell} \end{bmatrix} \xrightarrow{\mu} a$$

## Этап применения (test)

Алгоритм  $a$  для новых объектов  $x'_i$  выдаёт ответы  $a(x'_i)$



# Функционалы качества

$\mathcal{L}(a, x)$  — функция потерь (loss function). Величина ошибки алгоритма  $a \in A$  на объекте  $x \in X$ .

## Функции потерь для задач классификации

$\mathcal{L}(a, x) = [a(x) \neq y(x)]$  — индикатор ошибки

## Функции потерь для задач регрессии

- ▶  $\mathcal{L}(a, x) = |a(x) - y(x)|$  — абсолютное значение ошибки
- ▶  $\mathcal{L}(a, x) = (a(x) - y(x))^2$  — квадратичная ошибка

*Эмпирический риск* — функционал качества алгоритма  $a$  на  $X^\ell$ :

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}(a, x_i)$$

# Сведение задачи обучения к задаче оптимизации

Метод минимизации эмпирического риска

$$\mu(X^\ell) = \arg \min_{a \in A} Q(a, X^\ell)$$

## Пример

Метод наименьших квадратов: ( $Y = \mathbb{R}$ ,  $\mathcal{L}$  квадратична)

$$\mu(X^\ell) = \arg \min_{\theta} \sum_{i=1}^{\ell} (g(x_i, \theta) - y_i)^2$$

## Проблема обобщающей способности

- ▶ Найдём ли мы «закон природы» или переобучимся, то есть подгоним функцию  $g(x_i, \theta)$  под заданные точки?
- ▶ Будет ли  $a = \mu(X^\ell)$  приближать функцию  $y$  на всём  $X$ ?
- ▶ Будет ли  $Q(a, X^k)$  мало на новых данных — контрольной

# Демо: пример переобучения

Разбираемся с кодом!

# Переобучение

одна из главных проблем в машинном обучении

## Из-за чего возникает переобучение?

- ▶ избыточная сложность пространства параметров  $\Theta$ , лишние степени свободы в модели  $g(x, \theta)$  «тратятся» на чрезмерно точную подгонку под обучающую выборку
- ▶ переобучение есть всегда, когда есть оптимизация параметров по конечной (заведомо неполной) выборке

## Как обнаружить переобучение?

Эмпирически, путём разбиения выборки на train и test

## Нельзя избавиться от него совсем. Как минимизировать?

- ▶ минимизировать ошибку на валидации (HoldOut, Leave One Out, Cross Validation), но осторожно!
- ▶ накладывать ограничения на  $\theta$  (регуляризация)

# Эмпирические оценки обобщающей способности

- ▶ Эмпирический риск на тестовых данных (hold-out)

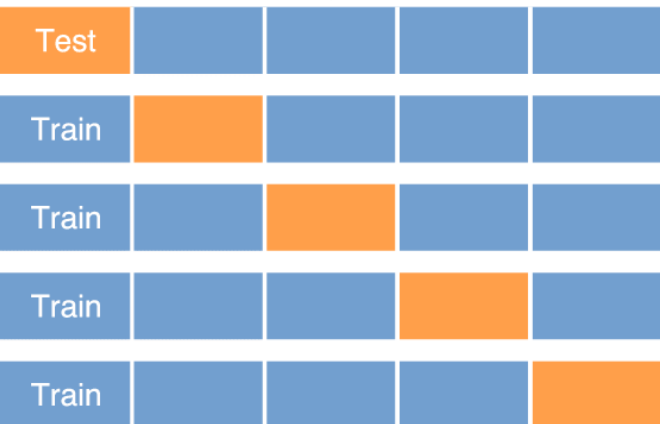
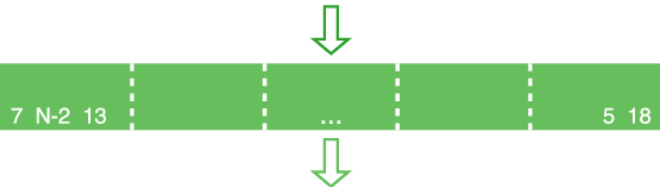
$$HO(\mu, X^\ell, X^k) = Q(\mu(X^\ell), X^k) \rightarrow \min$$

- ▶ Скользящий контроль (leave-one-out),  $L = \ell + 1$

$$LOO(\mu, X^\ell) = \frac{1}{L} \sum_{i=1}^L \mathcal{L}(\mu(X^\ell \setminus \{x_i\}), x_i) \rightarrow \min$$

- ▶ Кросс-проверка (cross-validation),  $L = \ell + k$ ,  $X^L = X_n^\ell \cup X_n^k$ :

$$CV(\mu, X^L) = \frac{1}{|N|} \sum_{n \in N} Q(\mu(X_n^\ell), X_n^k) \rightarrow \min$$



err<sub>1</sub>

err<sub>2</sub>

err<sub>3</sub>

err<sub>4</sub>

err<sub>5</sub>

mean  
error

# Ключевые события в машинном обучении

1997 IBM Deep Blue обыгрывает чемпиона мира по шахматам Гарри Каспарова

- ▶ 480 шахматных CPU
- ▶ перебор модификацией альфа-бета-отсечений
- ▶ две дебютные книги

# Ключевые события в машинном обучении

1997 IBM Deep Blue обыгрывает чемпиона мира по шахматам Гарри Каспарова

- ▶ 480 шахматных CPU
- ▶ перебор модификацией альфа-бета-отсечений
- ▶ две дебютные книги

2004 Соревнование беспилотных автомобилей: DARPA Grand Challenge

- ▶ призовой фонд \$1 млн
- ▶ в первом заезде победитель проехал 11.8 из 230 км



# Ключевые события в машинном обучении

1997 IBM Deep Blue обыгрывает чемпиона мира по шахматам Гарри Каспарова

- ▶ 480 шахматных CPU
- ▶ перебор модификацией альфа-бета-отсечений
- ▶ две дебютные книги

2004 Соревнование беспилотных автомобилей: DARPA Grand Challenge

- ▶ призовой фонд \$1 млн
- ▶ в первом заезде победитель проехал 11.8 из 230 км

2006 Запуск Google Translate

- ▶ сначала статистический машинный перевод
- ▶ мобильное приложение появилось в 2010

## 2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011-15 ImageNet: 25% -> 3.5% ошибок против 5% у людей

2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011-15 ImageNet: 25% -> 3.5% ошибок против 5% у людей

2015 Создание открытой компании OpenAI, Илон Маск и Сэм Альтман, обещали вложить \$1 млрд

2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011-15 ImageNet: 25% -> 3.5% ошибок против 5% у людей

2015 Создание открытой компании OpenAI, Илон Маск и Сэм Альтман, обещали вложить \$1 млрд

2016 Google DeepMind AlphaGo обыграл чемпиона мира по игре Го

2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011-15 ImageNet: 25% -> 3.5% ошибок против 5% у людей

2015 Создание открытой компании OpenAI, Илон Маск и Сэм Альтман, обещали вложить \$1 млрд

2016 Google DeepMind AlphaGo обыграл чемпиона мира по игре Го

2018 На аукционе Christie's картина, формально нарисованная ИИ, продана за 432 500\$

2011 40 лет развития DARPA CALO (Cognitive Assistant that Learns and Organizes)

- ▶ появление голосового помощника Apple Siri
- ▶ IBM Watson победил в телевизионной игре «Jeopardy!» (у нас «Своя игра»)

2011-15 ImageNet: 25% -> 3.5% ошибок против 5% у людей

2015 Создание открытой компании OpenAI, Илон Маск и Сэм Альтман, обещали вложить \$1 млрд

2016 Google DeepMind AlphaGo обыграл чемпиона мира по игре Го

2018 На аукционе Christie's картина, формально нарисованная ИИ, продана за 432 500\$

2020 AlphaFold 2 предсказывает структуру белков с точностью выше 90% для примерно двух третей белков в датасете

- ▶ Основные понятия машинного обучения:  
обучение по прецедентам (с учителем),  
объекты, признаки, ответы, модель алгоритмов, метод  
обучения, эмпирический риск, переобучение
- ▶ Проблема переобучения: HoldOut, LeaveOneOut,  
CrossValidation
- ▶ Ключевые события в машинном обучении