

• Введение в машинное обучение •
**Научный метод и основные понятия
машинного обучения**

Воронцов Константин Вячеславович

`k.v.vorontsov@phystech.edu`

`http://www.MachineLearning.ru/wiki?title=User:Vokov`

Этот курс доступен на странице вики-ресурса

`http://www.MachineLearning.ru/wiki`

«Введение в машинное обучение (курс лекций, К.В.Воронцов)»

МФТИ.ФПМИ.ИС.ИАД • 12 февраля 2026

«Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, искусственном интеллекте и **машинном обучении**» (2016)

Клаус Мартин Шваб,
президент Всемирного экономического форума



До этого про ML говорили лишь на научных конференциях; теперь AI/ML — это «новый двигатель прогресса», «технологии, которые меняют мир»...

«Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future»

- Цифровая и распределённая экономика
- Автоматизация и сокращение издержек
- Автономный транспорт и роботизация
- Оптимизация логистики и цепей поставок
- Оптимизация энергетических сетей (Energy Tech)
- Автоматизация банковских услуг (Fin Tech)
- Автоматизация юридических услуг (Legal Tech)
- Автоматизация образовательных услуг (Ed Tech)
- Автоматизация работы с кадрами (HR Tech)
- Персональная медицина (Med Tech)
- Автоматизация в сельском хозяйстве (Agro Tech)
- Автономные системы вооружений (Mil Tech)



- 1 Домашинная история машинного обучения**
 - Эмпирическая индукция и научный метод
 - Задача регрессии: исторические примеры
 - Задача классификации: исторические примеры
- 2 Базовые понятия и обозначения**
 - Табличные данные в задачах машинного обучения
 - Параметрические модели и алгоритмы обучения
 - Обучение и переобучение
- 3 Практика машинного обучения**
 - Задачи классификации, регрессии, ранжирования
 - Задачи со сложно структурированными данными
 - Философия машинного обучения

Принцип эмпирической индукции

«Не следует полагаться на сформулированные аксиомы и формальные базовые понятия, какими бы привлекательными и справедливыми они не казались. Законы природы нужно «расшифровывать» из фактов опыта.

Следует искать правильный метод анализа и обобщения опытных данных;

здесь логика Аристотеля не подходит в силу её абстрактности, оторванности от реальных процессов и явлений.»



Фрэнсис Бэкон
(1561–1626)

Таблица открытия: множество объектов $\{x_i: i = 1, \dots, \ell\}$

- $f_j(x)$ — измеряемые признаки объектов, $j = 1, \dots, n$
- $y_i \in \mathbb{R}$ — измеряемое значение целевого свойства x_i , либо $y_i \in \{0, 1\}$ — отсутствие или наличие целевого свойства

Фрэнсис Бэкон. Новый органон. 1620.

Восстановление зависимостей по эмпирическим данным

Дано: обучающая выборка объектов $x_i = (f_1(x_i), \dots, f_n(x_i)) \in X$ с ответами $y_i = y(x_i) \in Y$, $i = 1, \dots, \ell$

Найти: параметры w модели $a(x, w)$, приближающей зависимость $y: X \rightarrow Y$

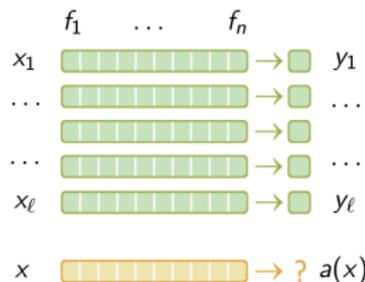
Критерий: минимум эмпирического риска

$$\sum_{i=1}^{\ell} \mathcal{L}(a(x_i, w), y_i) \rightarrow \min_w$$

$\mathcal{L}(a, y)$ — функция потерь (ответ модели a при правильном y)

Основные типы задач машинного обучения с учителем:

- $\mathcal{L}(a, y) = (a - y)^2$ в задачах регрессии, $Y = \mathbb{R}$
- $\mathcal{L}(a, y) = [a \neq y]$ в задачах классификации, $|Y| < \infty$



Метод наименьших квадратов (Гаусс, 1795)

Линейная модель регрессии:

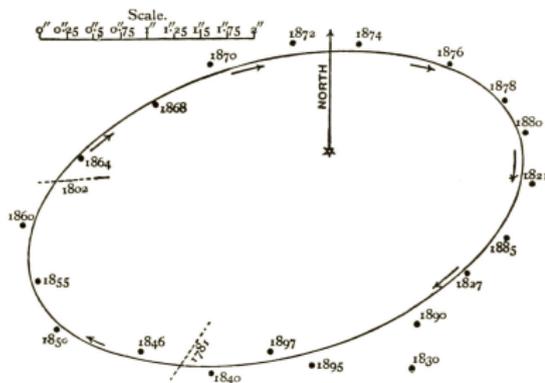
$$a(x, w) = \sum_{j=1}^n w_j f_j(x), \quad w \in \mathbb{R}^n$$

Метод наименьших квадратов:

$$\sum_{i=1}^{\ell} (a(x_i, w) - y_i)^2 \rightarrow \min_w$$



Карл Фридрих
Гаусс (1777–1855)



«Our principle, which we have made use of since 1795, has lately been published by Legendre...»

C.F. Gauss. Theory of the motion of the heavenly bodies moving about the Sun in conic sections. 1809.

Восстановление уравнения эллипса по точкам

Дано: $(x_i, y_i)_{i=1}^{\ell}$ — точки эллипса, в декартовых координатах, измеренные с погрешностями

Найти: параметры w_{ab} уравнения эллипса

$$w_{20}x^2 + w_{11}xy + w_{02}y^2 + w_{10}x + w_{01}y + w_{00} = 0$$

Критерий: метод наименьших квадратов (least squares)

$$\sum_{i=1}^{\ell} (w_{20}x_i^2 + w_{11}x_iy_i + w_{02}y_i^2 + w_{10}x_i + w_{01}y_i + w_{00})^2 \rightarrow \min_w$$

Это модель, линейная по параметрам $w_{20}, w_{11}, w_{02}, w_{10}, w_{01}, w_{00}$

Вопрос 1: почему такая нумерация параметров?

Вопрос 2: что есть признаки и что целевое свойство?

Вопрос 3: это обучение с учителем или без учителя?

Восстановление уравнения эллипса по точкам: 2-й способ

Дано: $(t_i, x_i, y_i)_{i=1}^{\ell}$ — точки эллипса, добавлены измерения моментов времени t_i , тоже с погрешностями

Найти: уравнение эллипса в параметрической форме

$$\begin{cases} x(t; R, a, b, x_0) = R \cos(at + b) + x_0 \\ y(t; r, a, b, y_0) = r \sin(at + b) + y_0 \end{cases}$$

Критерий: метод наименьших квадратов (least squares)

$$\sum_{i=1}^{\ell} (x(t_i; R, a, b, x_0) - x_i)^2 + (y(t_i; r, a, b, y_0) - y_i)^2 \rightarrow \min$$

Модель линейна по параметрам R, r, x_0, y_0 , не линейна по a, b

Вопрос 1: что есть признаки и что целевое свойство?

Вопрос 2: это обучение с учителем или без учителя?

Восстановление уравнения эллипса по точкам: 3-й способ

Дано: $(\varphi_i, \rho_i)_{i=1}^{\ell}$ — точки эллипса в полярных координатах

Найти: уравнение эллипса в полярных координатах

$$\rho(\varphi) = \frac{R}{1 - \varepsilon \cos \varphi}$$

Критерий: метод наименьших квадратов, два варианта:

$$\sum_{i=1}^{\ell} \left(\frac{R}{1 - \varepsilon \cos \varphi_i} - \rho_i \right)^2 \rightarrow \min_{R, \varepsilon}; \quad \sum_{i=1}^{\ell} (\varepsilon \rho_i \cos \varphi_i + R - \rho_i)^2 \rightarrow \min_{R, \varepsilon}$$

B1: что есть признаки? что целевое свойство?

B2: модель линейная или не линейная по параметрам?

B3: почему в модели два параметра, а не шесть?

B4: какие данные были у Гаусса? как он формализовал задачу?

B5: где фокус эллипса? как далее развивать модель?

История термина «регрессия» (Гальтон, 1886)

Дано: $(x_i, y_i)_{i=1}^{\ell}$ — отклонение роста отца (x_i) и взрослого сына (y_i) от среднего в популяции

Найти: линейную модель наследственности роста

$$a(x, w) = wx$$

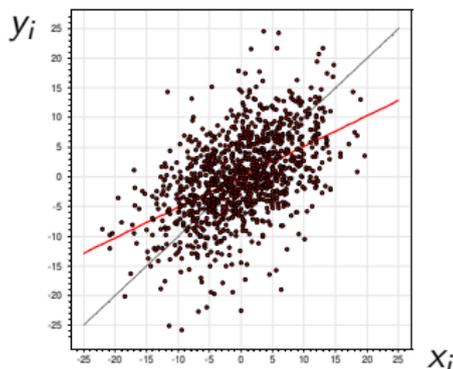
Критерий

наименьших квадратов:

$$\sum_{i=1}^{\ell} (wx_i - y_i)^2 \rightarrow \min_w$$

Решение:

$$w = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = 0.67$$



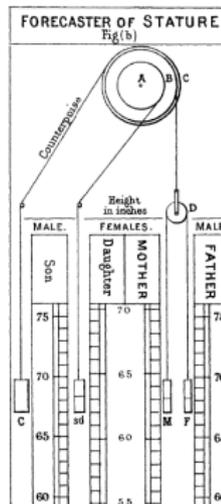
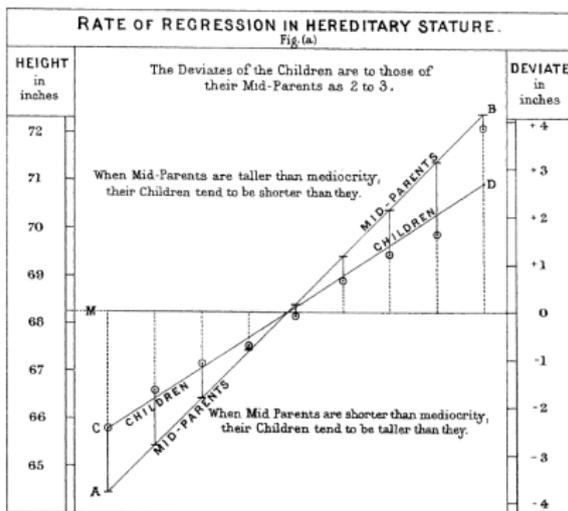
Фрэнсис
Гальтон
(1822–1911)

В: какой была бы популяция людей, если бы $w > 1$?

Galton F. Regression towards mediocrity in hereditary stature. 1886.

История термина «регрессия» (Гальтон, 1886)

«Регрессия к посредственности» — угол наклона меньше 1
Скрытый смысл: обратный ход исследования от данных к модели

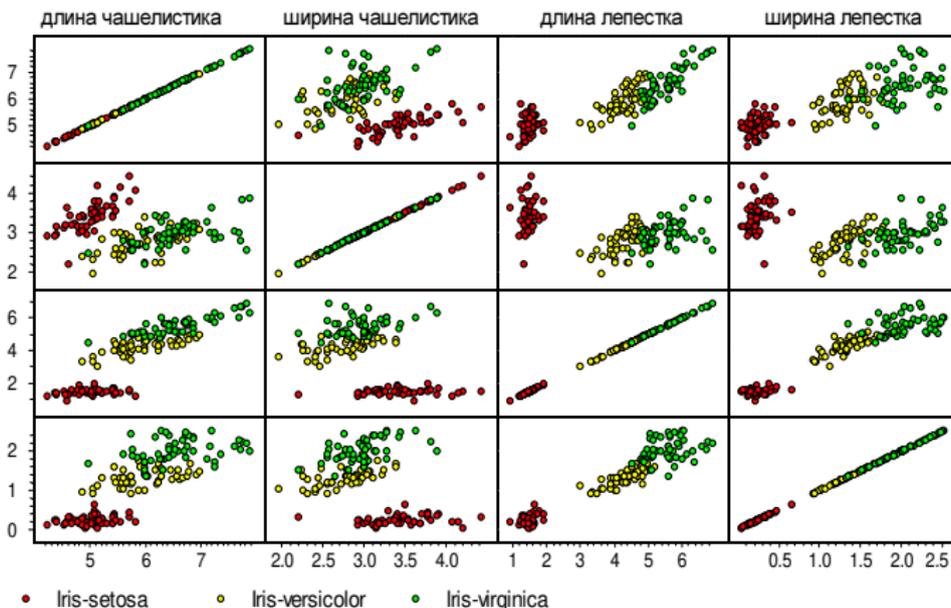


B: как учесть рост матери, дедушек, бабушек, дальних предков?

Galton F. Regression towards mediocrity in hereditary stature. 1886.

Задача классификации цветков ириса (Фишер, 1936)

Дано: $n = 4$ признака, $|Y| = 3$ класса, наблюдений $\ell = 150$



В: какую модель вы бы строили, исходя из графиков?

Линейный дискриминантный анализ (Фишер, 1936)

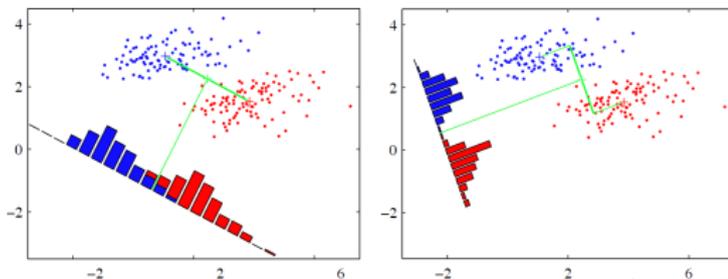
Найти линейную модель классификации:

$$a(x, w) = \text{sign} \left(\sum_{j=1}^n w_j f_j(x) - w_0 \right)$$

Критерий: в проекции на направляющий вектор w разделяющей гиперплоскости вероятность ошибки минимальна:



Рональд
Фишер
(1890–1962)



Fisher R. A. The use of multiple measurements in taxonomic problems. 1936.

Как задаются объекты. Векторное признаковое описание

$f_j: X \rightarrow D_j, j = 1, \dots, n$ — признаки объектов (features)

Скалярные (одномерные) типы признаков:

- $D_j = \{0, 1\}$ — *бинарный* признак f_j
- $|D_j| < \infty$ — *номинальный* признак f_j
- $|D_j| < \infty, D_j$ упорядочено — *порядковый* признак f_j
- $D_j = \mathbb{R}$ — *количественный* признак f_j

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

Табличные данные (feature data, матрица «объекты–признаки»)

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

Как задаются ответы. Типы задач

Задачи обучения с учителем (supervised learning):

на объектах $x_i \in X^\ell$ заданы правильные ответы $y_i = y(x_i)$

задачи классификации (classification, Y — class labels):

- $Y = \{0, 1\}$ или $\{-1, +1\}$ — на 2 класса (binary classification)
- $Y = \{1, \dots, M\}$ — на много классов (multiclass cl.)
- $Y = \{0, 1\}^M$ — на пересекающиеся классы (multilabel cl.)

задачи восстановления регрессии (regression):

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

задачи ранжирования (ranking, learning to rank):

- Y — конечное упорядоченное множество

Задачи обучения без учителя (unsupervised learning):

- ответов нет, требуется что-то делать с самими объектами

Как задаются предсказательные модели

Модель (predictive model) — параметрическое семейство функций

$$A = \{a(x, w) \mid w \in W\},$$

где $a: X \times W \rightarrow Y$ — фиксированная функция,

W — множество допустимых значений параметра w

Пример.

Линейная модель с вектором параметров $w = (w_1, \dots, w_n) \in \mathbb{R}^n$:

$$a(x, w) = \sum_{j=1}^n w_j f_j(x) \quad \text{— для регрессии и ранжирования, } Y = \mathbb{R}$$

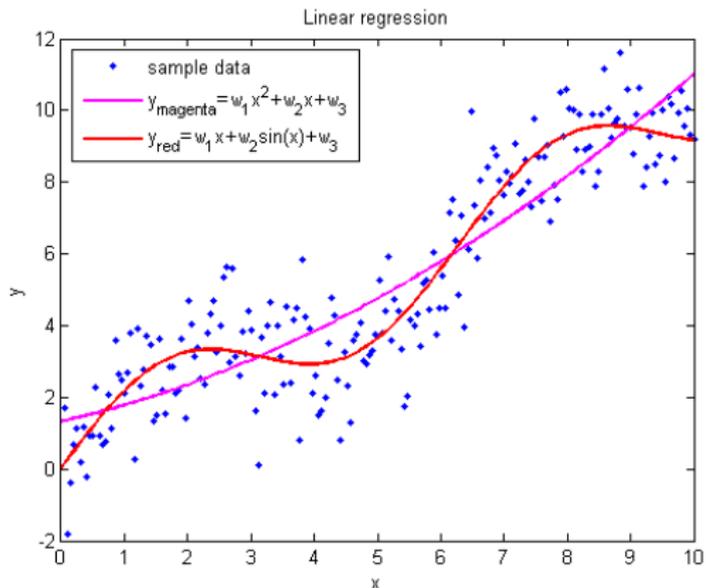
$$a(x, w) = \text{sign} \sum_{j=1}^n w_j f_j(x) \quad \text{— для классификации, } Y = \{-1, +1\}$$

V1: помните, как определяется функция $\text{sign}(z)$?

V2: как задать линейную модель, если классов много, $|Y| > 2$?

Пример: задача регрессии, синтетические данные

$X = Y = \mathbb{R}$, $\ell = 200$, $n = 3$ признака: $\{x, x^2, 1\}$ или $\{x, \sin x, 1\}$



- вычисление новых признаков может обогатить модель
- на практике очень важно взять адекватную модель

Алгоритм обучения, этапы обучения и применения

Этап обучения (train или fit):

алгоритм обучения (learning algorithm) $\mu: (X \times Y)^\ell \rightarrow W$
по выборке $X^\ell = (x_i, y_i)_{i=1}^\ell$ строит функцию $a(x, w)$,
оценивая (оптимизируя) **параметры модели $w \in W$** :

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

Этап применения (test или predict):

функция $a(x, w)$ для новых объектов x'_i выдаёт **ответы $a(x'_i, w)$** :

$$\begin{pmatrix} f_1(x'_1) & \dots & f_n(x'_1) \\ \dots & \dots & \dots \\ f_1(x'_k) & \dots & f_n(x'_k) \end{pmatrix} \xrightarrow{a} \begin{pmatrix} a(x'_1, w) \\ \dots \\ a(x'_k, w) \end{pmatrix}$$

Как задаются функции потерь

Функции потерь для задач классификации:

- $\mathcal{L}(w, x) = [a(x, w) \neq y(x)]$ — binary loss, индикатор ошибки
- $\mathcal{L}(w, x) = \exp(-b(x, w)y(x))$ — margin-based loss для моделей вида $a(x, w) = \text{sign } b(x, w)$ при $Y = \{-1, +1\}$, где $b(x, w)y(x)$ — отступ (margin) объекта x от $\{x: b(x, w) = 0\}$ — поверхности, разделяющей классы

Функции потерь для задач регрессии:

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

Метод наименьших квадратов — частный случай ERM:

$$\sum_{i=1}^{\ell} (a(x_i, w) - y_i)^2 \rightarrow \min_w$$

В: [ложь] = 0, [истина] = 1. Как называется эта нотация?

Пример Рунге. Аппроксимация функции полиномом

Функция $y(x) = \frac{1}{1 + 25x^2}$ на отрезке $x \in [-2, 2]$

Признаковое описание объекта $x \mapsto (1, x^1, x^2, \dots, x^n)$

Модель полиномиальной регрессии

$a(x, w) = w_0 + w_1x + \dots + w_nx^n$ — полином степени n

Обучение методом наименьших квадратов:

$$Q(w, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} (w_0 + w_1x_i + \dots + w_nx_i^n - y_i)^2 \rightarrow \min_{w_0, \dots, w_n}$$

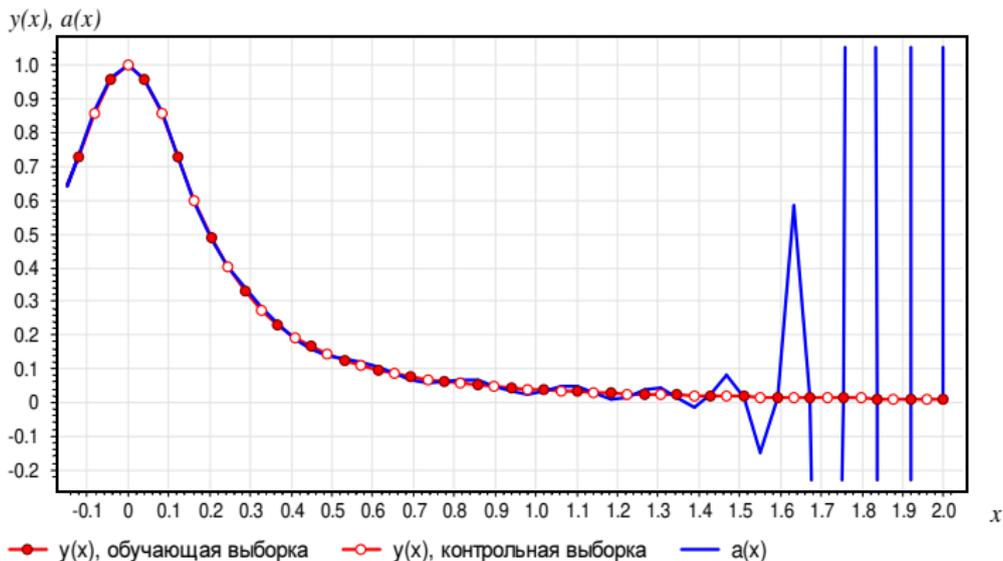
Обучающая выборка: $X^\ell = \{x_i = 4 \frac{i-1}{\ell-1} - 2 \mid i = 1, \dots, \ell\}$

Контрольная выборка: $X^k = \{x_i = 4 \frac{i-0.5}{\ell-1} - 2 \mid i = 1, \dots, \ell - 1\}$

Что происходит с $Q(w, X^\ell)$ и $Q(w, X^k)$ при увеличении n ?

Пример Рунге. Переобучение при $n = 38$, $\ell = 50$

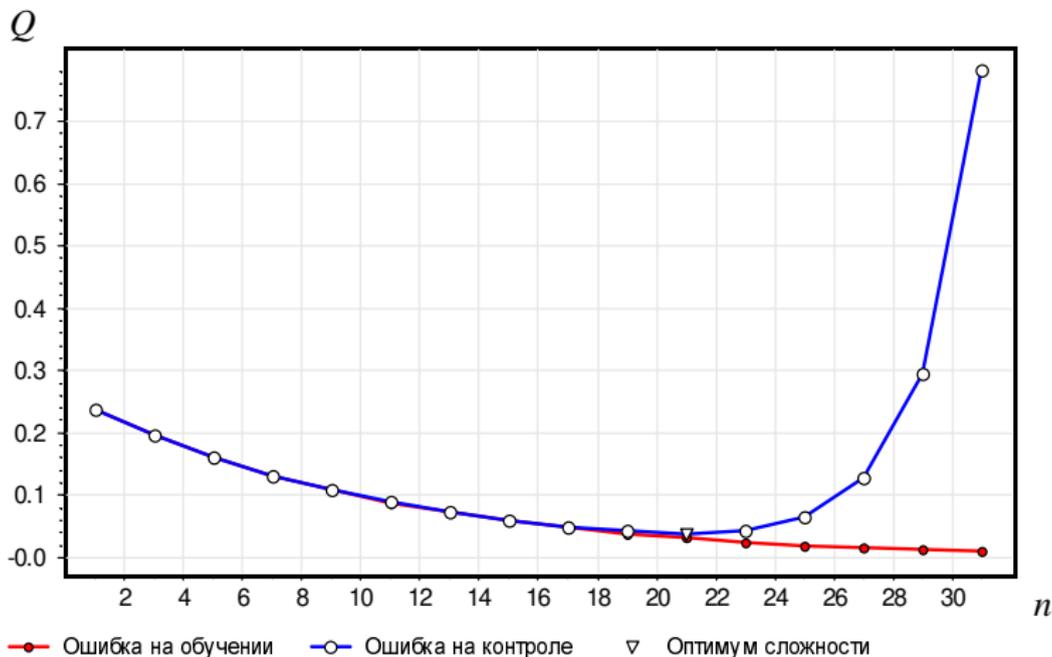
$$y(x) = \frac{1}{1 + 25x^2}; \quad a(x) \text{ — полином степени } n = 38$$



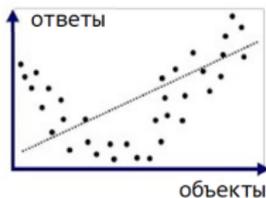
В: что не так на концах отрезка? как это интерпретировать?

Пример Рунге. Зависимость Q от степени полинома n

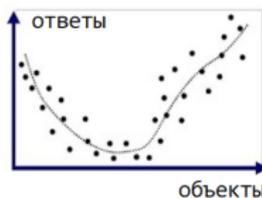
Переобучение — это когда $Q(\mu(X^{\ell}), X^k) \gg Q(\mu(X^{\ell}), X^{\ell})$:



Проблемы недообучения и переобучения

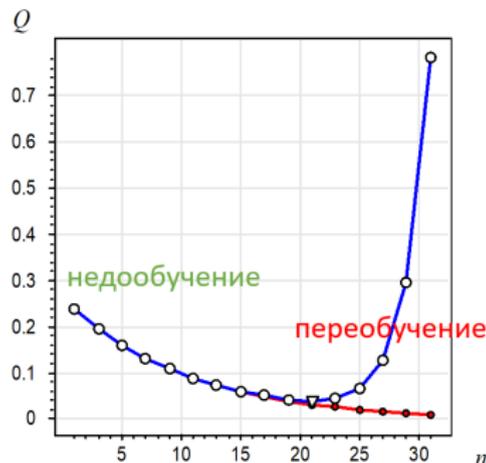


недообучение



переобучение

- **Недообучение** (underfitting):
данных много,
параметров недостаточно,
модель простая, негибкая
- **Переобучение** (overfitting):
данных мало, параметров
слишком много, модель
сложная, избыточно гибкая



Переобучение — ключевая проблема в машинном обучении

- 1 Из-за чего возникает переобучение?**
 - избыточные параметры в модели $a(x, w)$ «расходятся» на чрезмерно точную подгонку под обучающую выборку
 - выбор a из A производится по неполной информации X^ℓ
- 2 Как обнаружить переобучение?**
 - эмпирически, путём разбиения выборки на **train** и **test** (на **test** должны быть известны правильные ответы)
- 3 Избавиться от него нельзя. Как его минимизировать?**
 - увеличивать объём обучающих данных (big data)
 - накладывать ограничения на w (регуляризация)
 - минимизировать одну из теоретических оценок
 - выбирать лучшую модель (model selection) по оценкам обобщающей способности (generalization performance)

Эмпирические оценки обобщающей способности

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

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

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

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

- Кросс-проверка (cross-validation), $L = \ell + k$:

$$\text{CV}(\mu, X^L) = \frac{1}{|P|} \sum_{p \in P} Q(\mu(X_p^\ell), X_p^k) \rightarrow \min$$

где P — множество разбиений $X^L = X_p^\ell \sqcup X_p^k$, $p \in P$

Задачи медицинской диагностики

Объект — пациент в определённый момент времени.

Классы: диагноз или способ лечения или исход заболевания.

Примеры признаков:

- **бинарные:** пол, головная боль, слабость, тошнота, и т. д.
- **порядковые:** тяжесть состояния, желтушность, и т. д.
- **количественные:** возраст, пульс, артериальное давление, содержание гемоглобина в крови, доза препарата, и т. д.

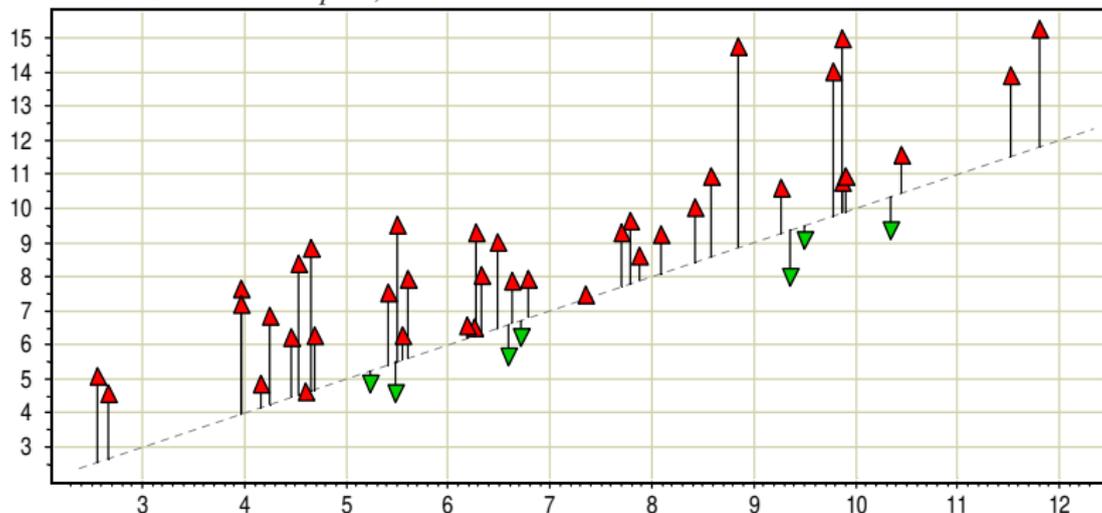
Особенности задачи:

- обычно много «пропусков» в данных;
- нужна интерпретируемая модель классификации;
- нужно выделять *синдромы* — сочетания *симптомов*;
- нужна оценка вероятности исхода.

Задача медицинской диагностики. Пример переобучения

Задача предсказания отдалённого результата хирургического лечения атеросклероза; точки — различные решающие правила

Частота ошибок на контроле, %



Частота ошибок на обучении, %

Задачи распознавания месторождений

Объект — геологический район (рудное поле).

Классы — есть или нет полезное ископаемое.

Примеры признаков:

- **бинарные:** присутствие крупных зон смятия и рассланцевания, и т. д.
- **порядковые:** минеральное разнообразие; мнения экспертов о наличии полезного ископаемого, и т. д.
- **количественные:** содержания сурьмы, присутствие в рудах антимонита, и т. д.

Особенности задачи:

- проблема «малых данных» — для редких типов месторождений объектов много меньше, чем признаков.

Задача кредитного скоринга

Объект — заявка на выдачу банком кредита.

Классы — bad или good.

Примеры признаков:

- бинарные: пол, наличие телефона, и т. д.
- номинальные: место проживания, профессия, работодатель, и т. д.
- порядковые: образование, должность, и т. д.
- количественные: возраст, зарплата, стаж работы, доход семьи, сумма кредита, и т. д.

Особенности задачи:

- нужно оценивать вероятность дефолта $P(y(x) = \text{bad})$.

Задача категоризации текстовых документов

Объект — текстовый документ.

Классы — рубрики иерархического тематического каталога.

Примеры признаков:

- **номинальные:** автор, издание, год, и т. д.
- **количественные:** для каждого термина — частота в тексте, в заголовках, в аннотации, и т. д.

Особенности задачи:

- лишь небольшая часть документов имеют метки y_i ;
- документ может относиться к нескольким рубрикам;
- в каждом ребре дерева свой классификатор на 2 класса.

Задача прогнозирования стоимости недвижимости

Объект — квартира в Москве.

Ответ — оценка стоимости, числовая величина.

Примеры признаков:

- **бинарные:** балкон, лифт, мусоропровод, охрана, и т. д.
- **номинальные:** район города, тип дома (кирпичный/панельный/блочный/монолит), и т. д.
- **количественные:** число комнат, площадь, этаж, расстояние до центра, до метро, возраст дома, и т. д.

Особенности задачи:

- выборка неоднородна, стоимость меняется со временем;
- разнотипные признаки;
- для линейной модели нужны преобразования признаков;

Задача ранжирования поисковой выдачи

Объект — пара (короткий текстовый запрос, документ).

Ответ — релевантность, числовая оценка.

Классы в обучающих данных — релевантен / не релевантен,
разметка делается людьми — ассессорами.

Примеры количественных признаков:

- частота слов запроса в документе,
- число ссылок на документ,
- число кликов на документ: всего, по данному запросу.

Особенности задачи:

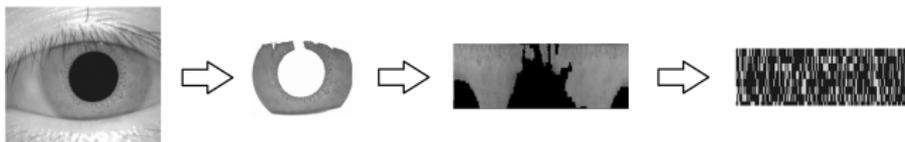
- сверхбольшие выборки документов;
- оптимизируется не число ошибок, а качество ранжирования;
- проблема конструирования признаков по сырым данным.

Задачи биометрической идентификации личности

Идентификация личности по отпечаткам пальцев



Идентификация личности по радужной оболочке глаза



Особенности задач:

- нетривиальная предобработка для извлечения признаков
- высочайшие требования к точности

J. Daugman. High confidence visual recognition of persons by a test of statistical independence. 1993

Входные данные сложной структуры

Сложные (не скалярные) типы признаков:

- текст, символьная последовательность
- сигнал, числовая последовательность
- чёрно-белое, серое изображение — 2D-матрица
- цветное, многозональное изображение — 3D-матрица
- видео, последовательность изображений — 4D-матрица
- транзакции, взаимодействия объектов друг с другом
- всё это вместе — мультимодальные данные

Выделение/генерация признаков (feature extraction/generation)

- вычисление признаков по формулам (feature engineering)
- обучаемая генерация вектора признаков (feature learning):

$f(x, w')$, $f: X \times W' \rightarrow \mathbb{R}^n$ — модель векторизации объекта

$$\sum_{i=1}^{\ell} \mathcal{L}(w, f(x_i, w')) \rightarrow \min_{w, w'} \quad \text{— обучение векторизатора}$$

Выходные данные сложной структуры. Генеративные модели

- **Статистический машинный перевод:**
объект — предложение на естественном языке
ответ — его перевод на другой язык
- **Перевод речи в текст:**
объект — аудиозапись речи человека
ответ — текстовая запись речи
- **Большая языковая модель, LLM:**
объект — текстовый контекст + запрос пользователя
ответ на запрос (промпт) пользователя
- **Генерация изображений по описанию:**
объект — текст с описанием изображения/видео
ответ — изображение/видео
- **Управление беспилотным аппаратом:**
объект — поток данных с видеокамер, датчиков
ответ — поток решений и управляющих сигналов

Научный метод познания: основные шаги и принципы

Наблюдения (эмпирический опыт, эксперименты, измерения)

Гипотеза (модель, теория) объясняет и обобщает наблюдения

- *Принцип верифицируемости* (Фрэнсис Бэкон): гипотеза подтверждается измеримыми наблюдениями
- *Принцип фальсифицируемости* (Карл Поппер): должны существовать способы опровергнуть гипотезу
- *Принцип соответствия* (Нильс Бор): новая гипотеза или теория должна включать прежнюю как частный случай
- *Принцип минимальной достаточности* («брита Оккама»): среди всех объяснений следует выбирать самое простое
- *Принцип воспроизводимости* (Роберт Бойль): открыто предоставлять всё необходимое для повторения результата
- *Принцип научной честности* (Ричард Фейнман): открыто дискутировать возможные опровержения гипотезы, «слабые места», противоречия, границы применимости

Машинное обучение как автоматизация научного метода

Наблюдения, измерения → выборка данных

Гипотеза → модель, параметрическое семейство функций

- *Принцип верифицируемости* (Фрэнсис Бэкон):
→ обучение (train) путём оптимизации параметров модели
- *Принцип фальсифицируемости* (Карл Поппер):
→ проверка (test) обученной модели на новых данных
- *Принцип соответствия* (Нильс Бор):
→ эксперименты с постепенным усложнением модели
- *Принцип минимальной достаточности* («бритва Оккама»):
→ своевременное прекращение усложнений
- *Принцип воспроизводимости* (Роберт Бойль):
→ культура открытых данных и открытого кода
- *Принцип научной честности* (Ричард Фейнман):
→ открытое тестирование моделей на бенчмарках,
сравнение своей модели с SOTA (State-Of-The-Art)

Машинное обучение — автоматизация научного метода

Задача машинного обучения начинается с её постановки, **ДНК** задачи (**Д**ано, **Н**айти, **К**ритерий)

Основные понятия машинного обучения:

- объект, ответ, признак, модель, функция потерь
- эмпирический риск, алгоритм обучения, переобучение

Три главных принципа машинного обучения:

- эмпирическая индукция Фрэнсиса Бэкона
- минимизация (+регуляризация) эмпирического риска
- обучаемая векторизация данных (Deep Learning)

Три основных типа прикладных задач:

- вектор $[\rightarrow$ вектор] \rightarrow скаляр (табличные входные данные)
- структура \rightarrow вектор \rightarrow скаляр (сложные данные на входе)
- структура \rightarrow вектор \rightarrow структура (... и на выходе тоже)