Методы машинного обучения (ММО)
Вводная лекция

Воронцов Константин Вячеславович
www.MachineLearning.ru/wiki?title=User:Vokov
вопросы к лектору: k.vorontsov@iai.msu.ru

материалы курса:
github.com/MSU-ML-COURSE/ML-COURSE-25-26
орг.вопросы по курсу: ml.cmc@mail.ru

ВМК МГУ • 2 сентября 2025

Содержание

Оптимизационная постановка задач машинного обучения

X — пространство *объектов*
Y — множество *ответов* (предсказаний / оценок / прогнозов)
$y(x)$, $y\colon X \to Y$ — неизвестная зависимость (target function)

Дано:
$\{x_1, \ldots, x_\ell\} \subset X$ — *обучающая выборка* (training sample)
$a(x, w)$, $a\colon X \times W \to Y$ — параметрическая модель зависимости

Найти:
$w \in W$ — вектор параметров модели такой, что $a(x, w) \approx y(x)$

Критерий — минимум эмпирического риска:

$$\sum_{i=1}^{\ell} \mathscr{L}(w, x_i) \to \min_w \quad \text{(empirical risk minimization, ERM)}$$

где $\mathscr{L}(w, x)$ — *функция потерь* (loss function) — тем больше, чем сильнее $a(x, w)$ отклоняется от правильного ответа $y(x)$

Как задаются объекты. Векторное признаковое описание

$f_j : X \to D_j$, $j = 1, \ldots, n$ — 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 $\big(f_1(x), \ldots, f_n(x)\big)$ — *признаковое описание* объекта x

Матрица «*объекты–признаки*» (feature data)

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

Как задаются объекты. Данные сложной структуры

Сложные типы признаков:

- текст, символьная дискретнозначная последовательность
- сигнал, непрерывнозначная 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 признаков (feature learning):
  $f(x, w')$, $f : X \times W' \to \mathbb{R}^n$ — модель векторизации объекта
  $\sum_{i=1}^{\ell} \mathscr{L}\big(w, f(x_i, w')\big) \to \min_{w,w'}$ — обучение векторизатора

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)
- $Y = \{1, \ldots, M\}$ — на много классов (multiclass c.)
- $Y = \{0, 1\}^M$ — на пересекающиеся классы (multilabel c.)

задачи ~~восстановления~~ регрессии (regression):

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

задачи ранжирования (ranking, learning to rank):

- Y — конечное упорядоченное множество

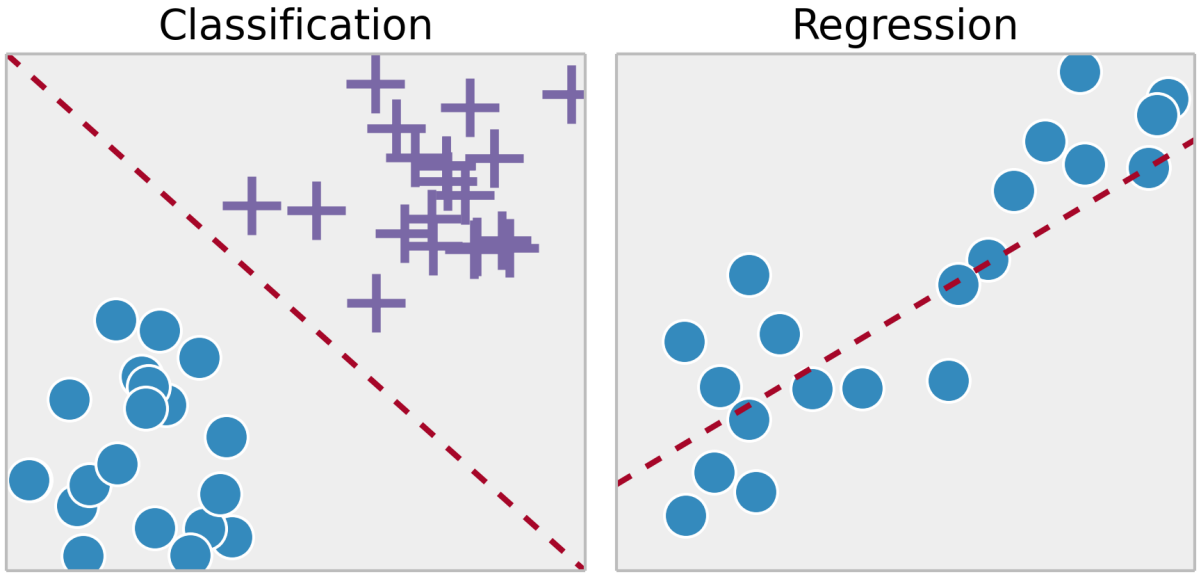
Задачи обучения без учителя (unsupervised learning):

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

Статистическое (машинное) обучение с учителем

= обучение по прецедентам
= восстановление зависимостей по эмпирическим данным
= предсказательное моделирование
= аппроксимация функций по заданным точкам

Два основных типа задач — *классификация* и *регрессия*



Как задаются предсказательные модели

Модель (predictive model) — параметрическое семейство функций

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

где $a\colon X \times W \to Y$ — фиксированная функция,
W — множество допустимых значений параметра w

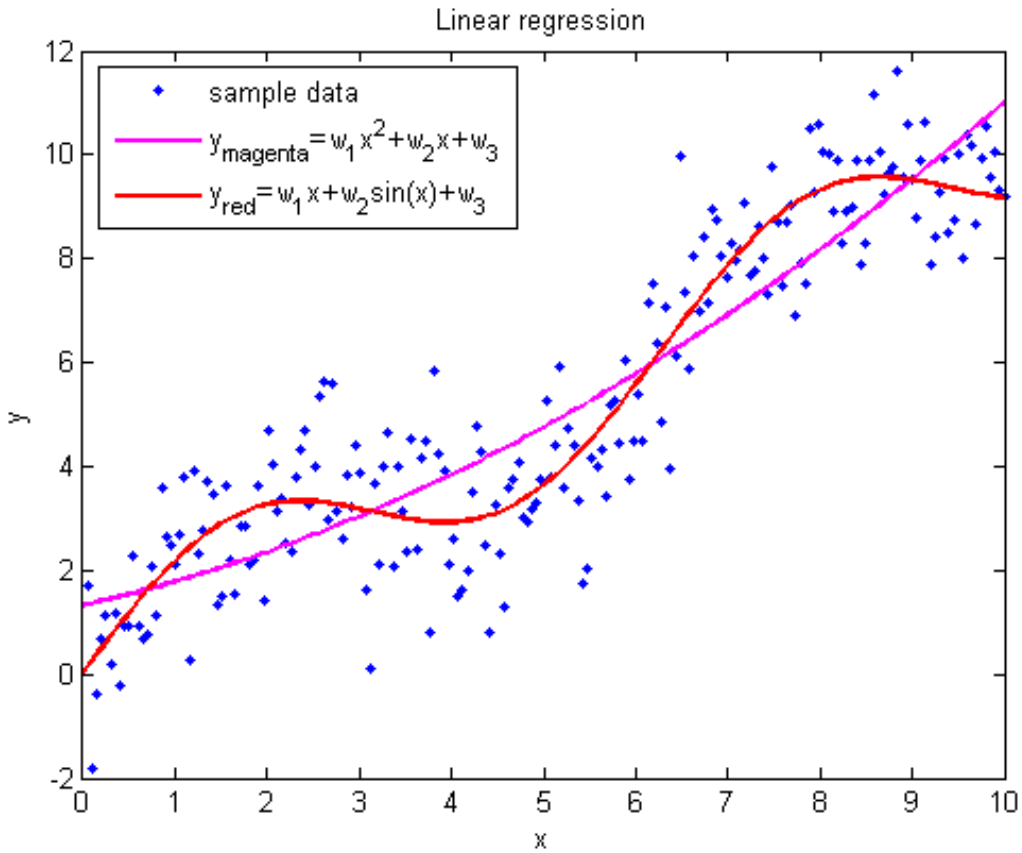
Пример.
Линейная модель с вектором параметров $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$:

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

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

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
- важно подобрать как можно более адекватную модель

Алгоритм обучения, этапы обучения и применения

Этап *обучения* (train):

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

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

Этап *применения* (test):

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:

- $\mathscr{L}(w,x) = \big[a(x,w) \neq y(x)\big]$ — индикатор ошибки
- для модели $a(x,w) = \operatorname{sign} b(x,w)$, $Y = \{-1,+1\}$:
  $\mathscr{L}(w,x) = L\big(b(x,w)y(x)\big)$ — margin-based функция потерь,
  $L(M)$ — непрерывная невозрастающая функция от
  $M(x,w) = b(x,w)y(x)$ — *отступ (margin)* объекта x

Функции потерь для задач регрессии:

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

Метод наименьших квадратов — частный случай ERM:

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

Пример Рунге. Аппроксимация функции полиномом

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

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

Модель полиномиальной регрессии

$$a(x, w) = w_0 + w_1 x + \cdots + w_n x^n \quad \text{— полином степени } n$$

Обучение методом наименьших квадратов:

$$Q(w, X^\ell) = \sum_{i=1}^{\ell} (w_0 + w_1 x_i + \cdots + w_n x_i^n - y_i)^2 \to \min_{w_0, \ldots, w_n}$$

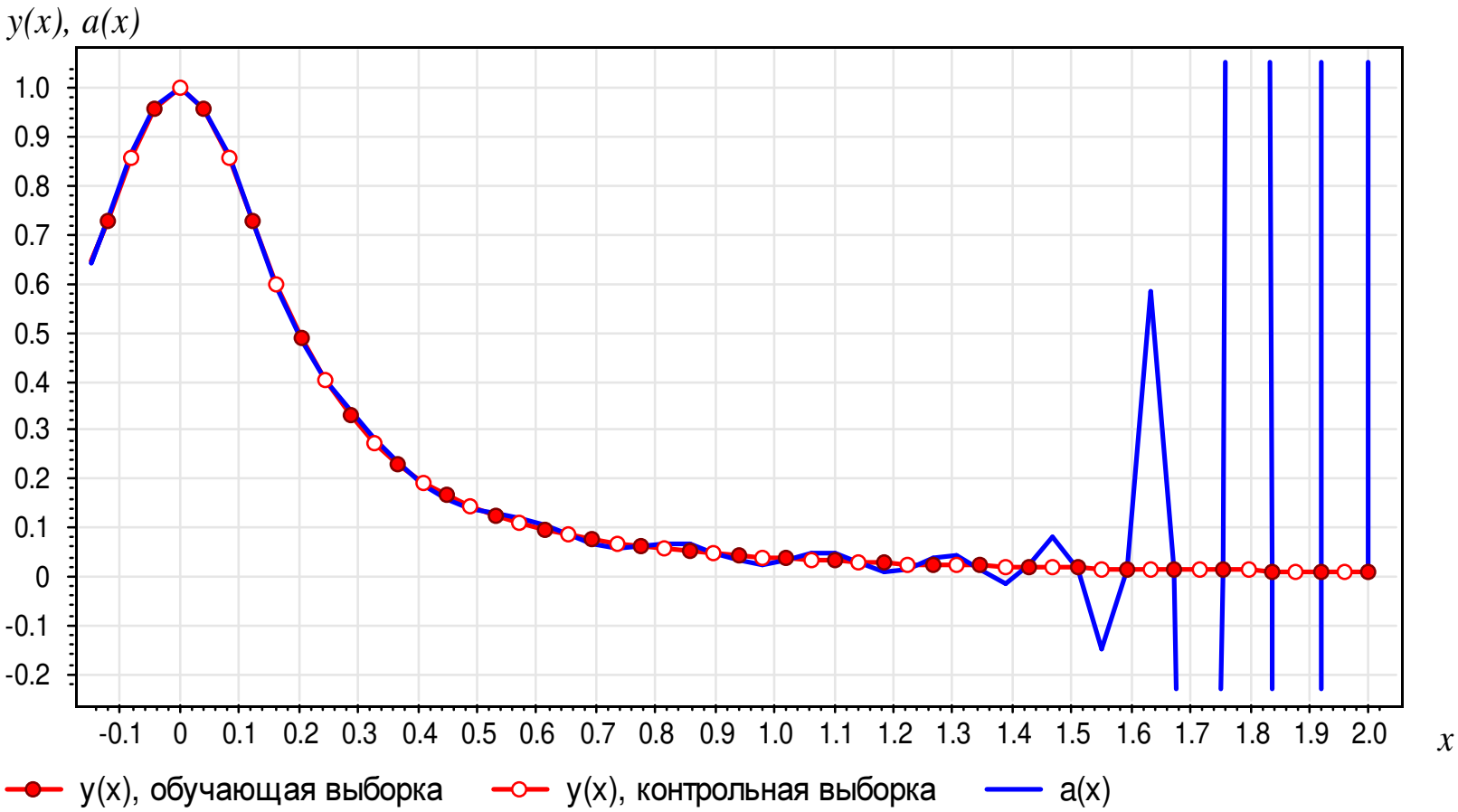
Обучающая выборка: $X^\ell = \left\{ x_i = 4\frac{i-1}{\ell-1} - 2 \;\middle|\; i = 1, \ldots, \ell \right\}$

Контрольная выборка: $X^k = \left\{ x_i = 4\frac{i-0.5}{\ell-1} - 2 \;\middle|\; i = 1, \ldots, \ell - 1 \right\}$

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

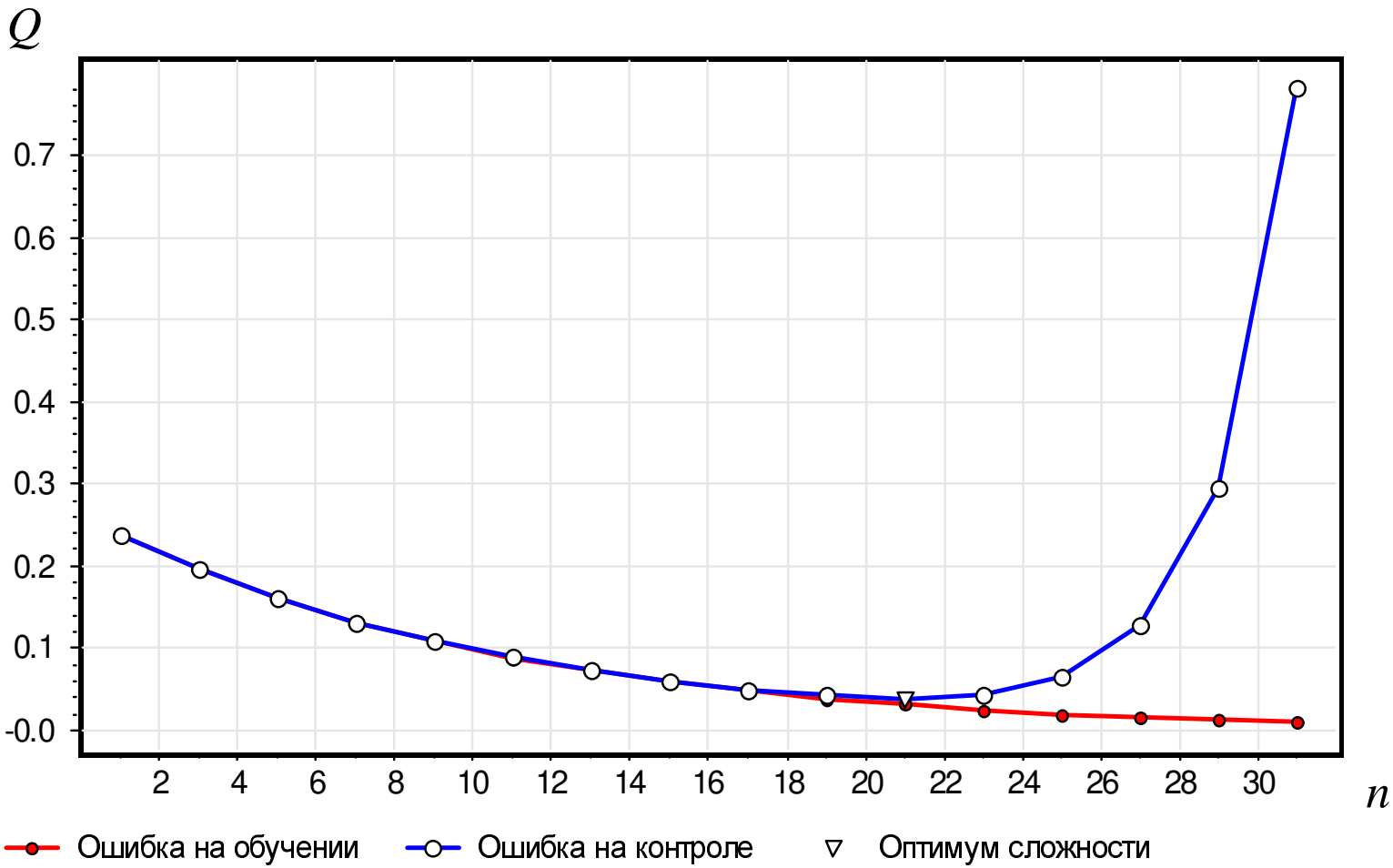
Пример Рунге. Переобучение при $n = 38$, $\ell = 50$

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

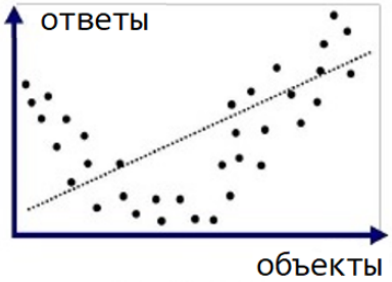


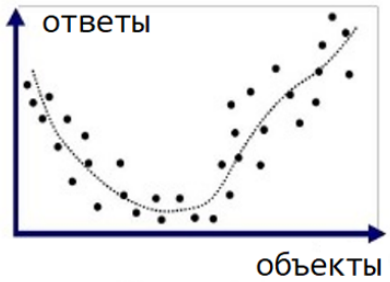
Пример Рунге. Зависимость Q от степени полинома n

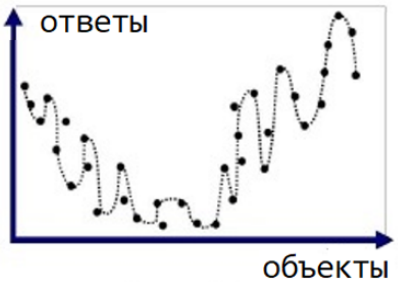
Переобучение — это когда $Q\big(\mu(X^\ell), X^k\big) \gg Q\big(\mu(X^\ell), X^\ell\big)$:



Проблемы недообучения и переобучения







недообучение

переобучение

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

Переобучение — ключевая проблема в машинном обучении

1. **Из-за чего возникает переобучение?**
   — избыточные параметры в модели $a(x, w)$ «расходуются» на чрезмерно точную подгонку под обучающие данные
   — 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):

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

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

$$\mathrm{LOO}(\mu, X^L) = \frac{1}{L} \sum_{i=1}^{L} \mathscr{L}(\mu(X^L \setminus \{x_i\}), x_i) \to \min$$

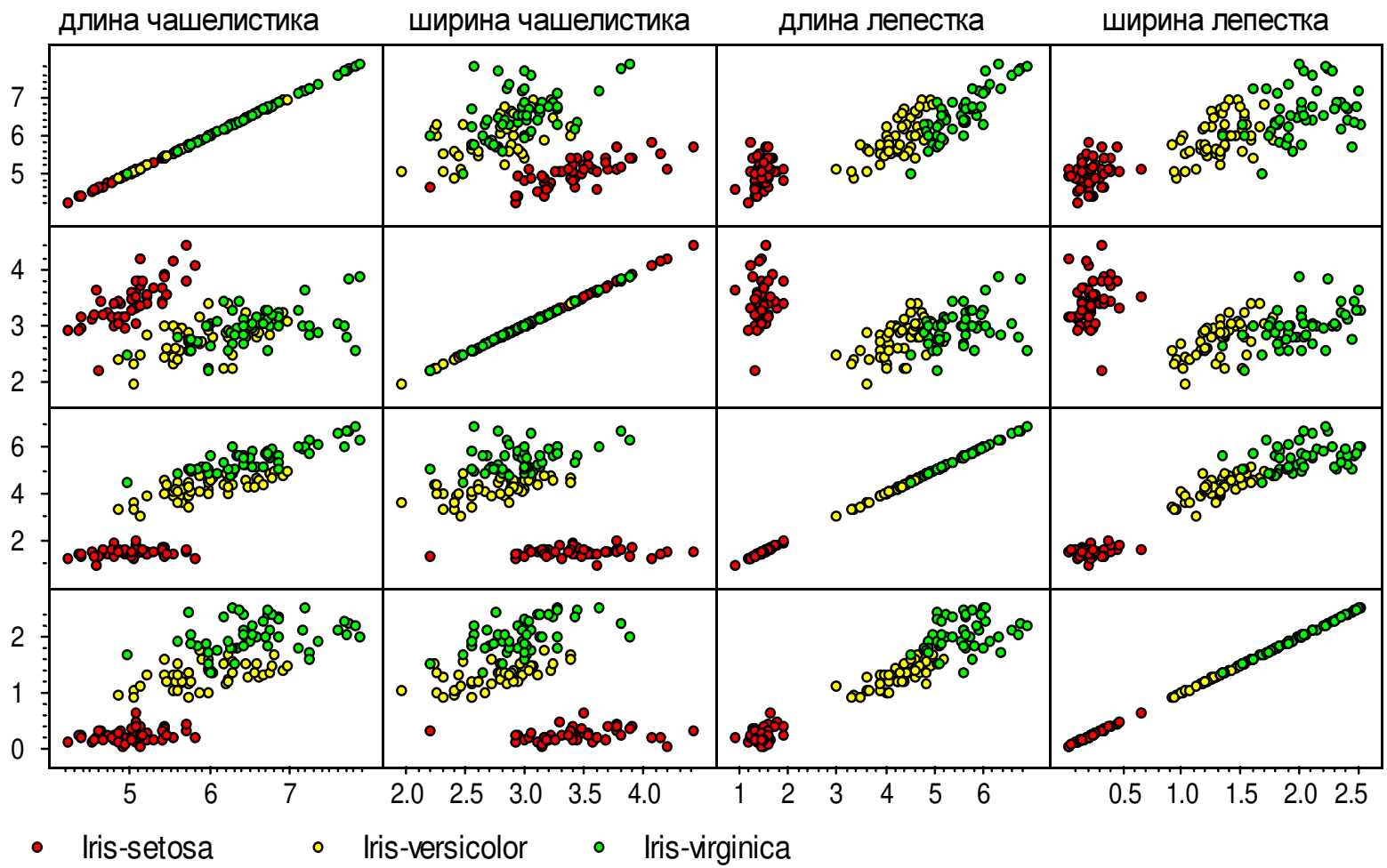
- Кросс-проверка (cross-validation), $L = \ell + k$:

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

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

Задача классификации цветков ириса (Фишер, 1936)

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

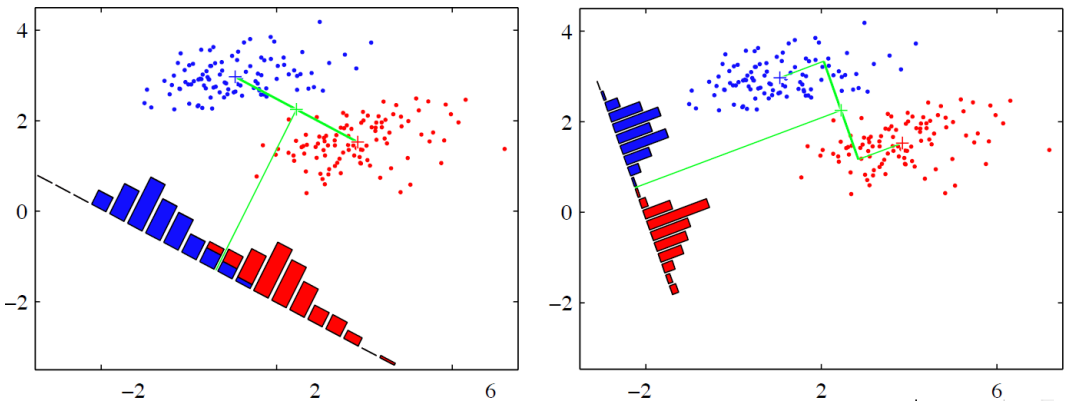


Линейный дискриминантный анализ (Фишер, 1936)

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.

Задачи медицинской диагностики

Объект — пациент в определённый момент времени.

Классы: диагноз или способ лечения или исход заболевания.

Примеры признаков:

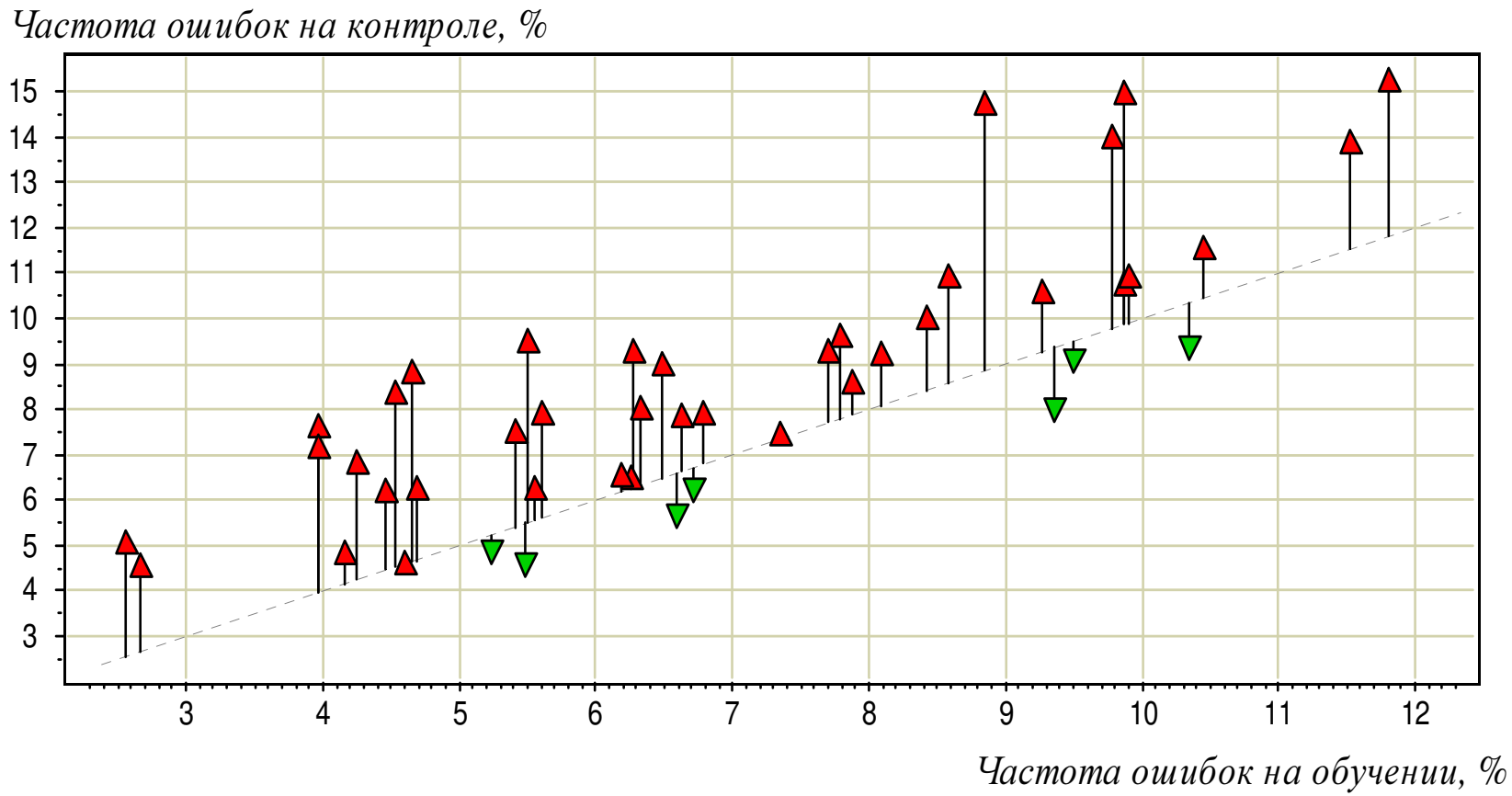
- **бинарные:** пол, головная боль, слабость, тошнота, и т. д.
- **порядковые:** тяжесть состояния, желтушность, и т. д.
- **количественные:** возраст, пульс, артериальное давление, содержание гемоглобина в крови, доза препарата, и т. д.

Особенности задачи:

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

Задача медицинской диагностики. Пример переобучения

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



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 $P\big(y(x) = \text{bad}\big)$.

Задача предсказания оттока клиентов

Объект — абонент в определённый момент времени.

Классы — уйдёт или не уйдёт в следующем месяце.

Примеры признаков:

- **бинарные:** корпоративный клиент, включение услуг, и т. д.
- **номинальные:** тарифный план, регион проживания, и т. д.
- **количественные:** длительность разговоров (входящих, исходящих, СМС, и т. д.), частота оплаты, и т. д.

Особенности задачи:

- нужно оценивать вероятность ухода;
- сверхбольшие выборки;
- признаки приходится вычислять по «сырым» данным.

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.

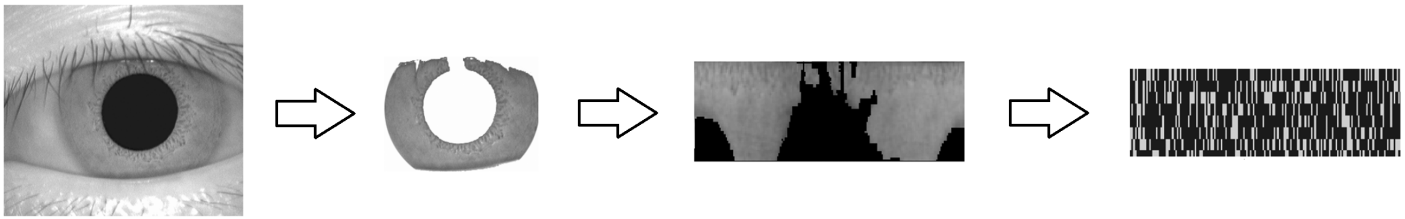
Особенности задачи:

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

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

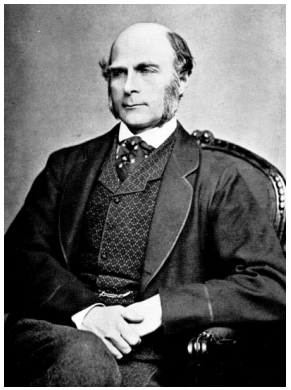
J. Daugman. High confidence visual recognition of persons by a test of statistical independence. 1993

История термина «регрессия» (Гальтон, 1886)

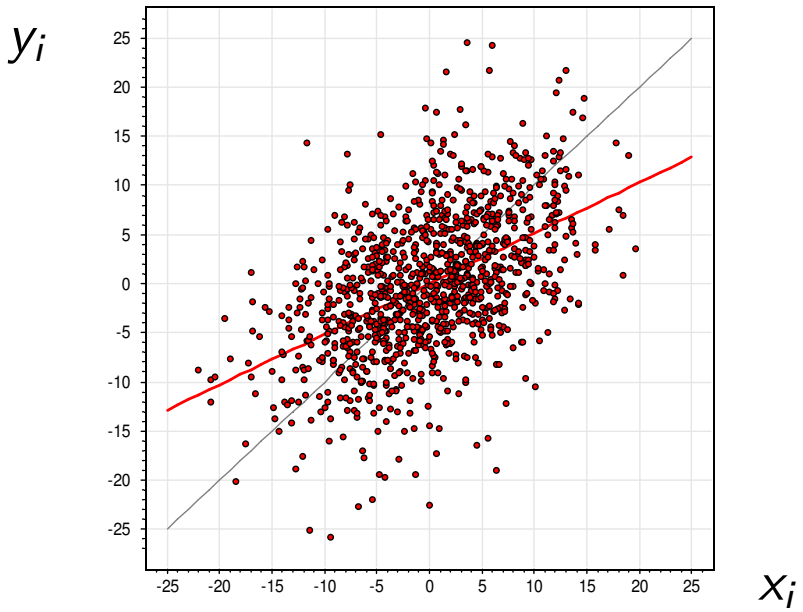
Дано: $(x_i, y_i)_{i=1}^{\ell}$ — отклонение роста отца (x_i) и взрослого сына (y_i) от среднего в популяции

Найти: модель наследственности роста $y(x) = kx$

Критерий: метод наименьших квадратов



Фрэнсис Гальтон (1822–1911)

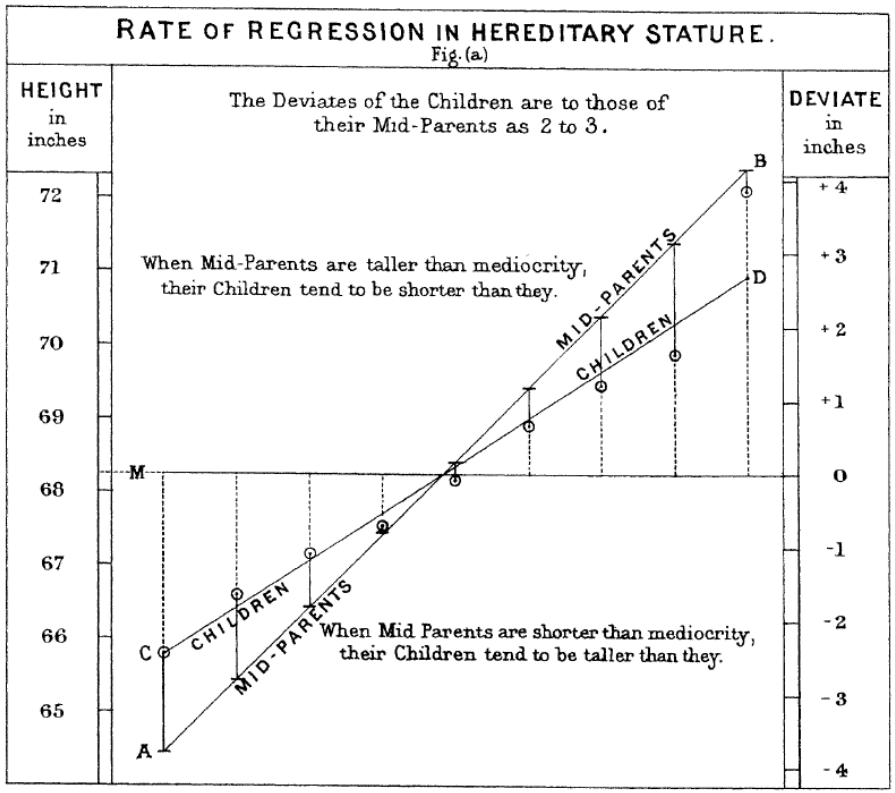


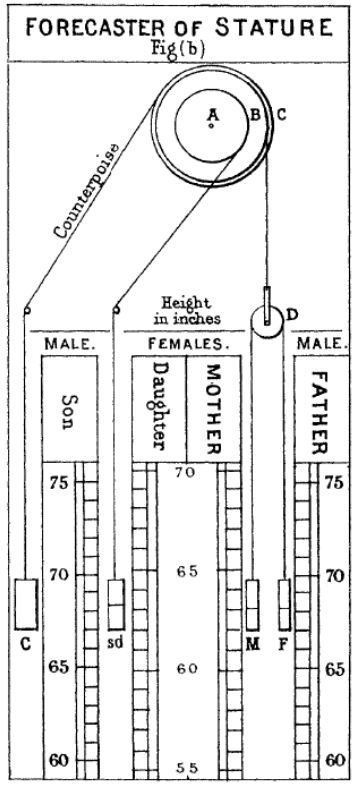
Galton F. Regression towards mediocrity in hereditary stature. 1886.

История термина «регрессия» (Гальтон, 1886)

«Регрессия к посредственности» — угол наклона меньше 1

Скрытый смысл: обратный ход исследования от данных к модели





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

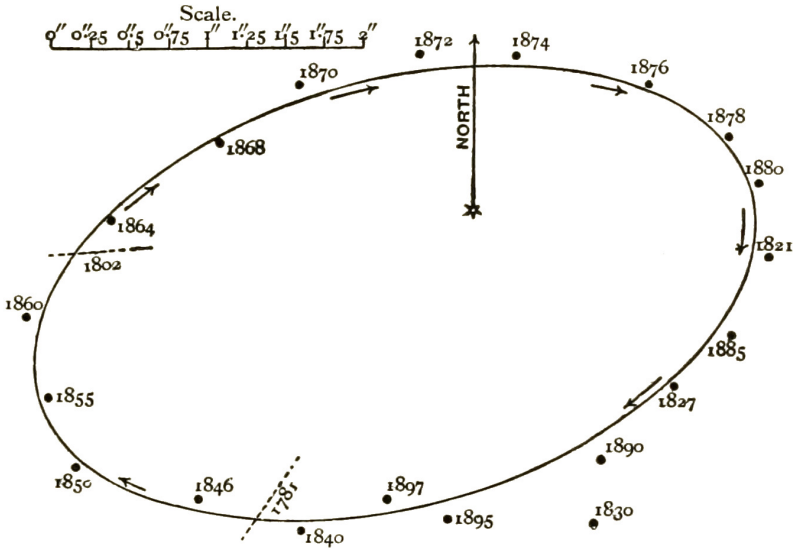
Метод наименьших квадратов (Гаусс, 1795)

Линейная модель регрессии:

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

Метод наименьших квадратов:

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



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.

Задача прогнозирования стоимости недвижимости

Объект — квартира в Москве.

Примеры признаков:

- **бинарные:** наличие балкона, лифта, мусоропровода, охраны, и т. д.
- **номинальные:** район города, тип дома (кирпичный/панельный/блочный/монолит), и т. д.
- **количественные:** число комнат, жилая площадь, 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 прогнозирования объёмов продаж

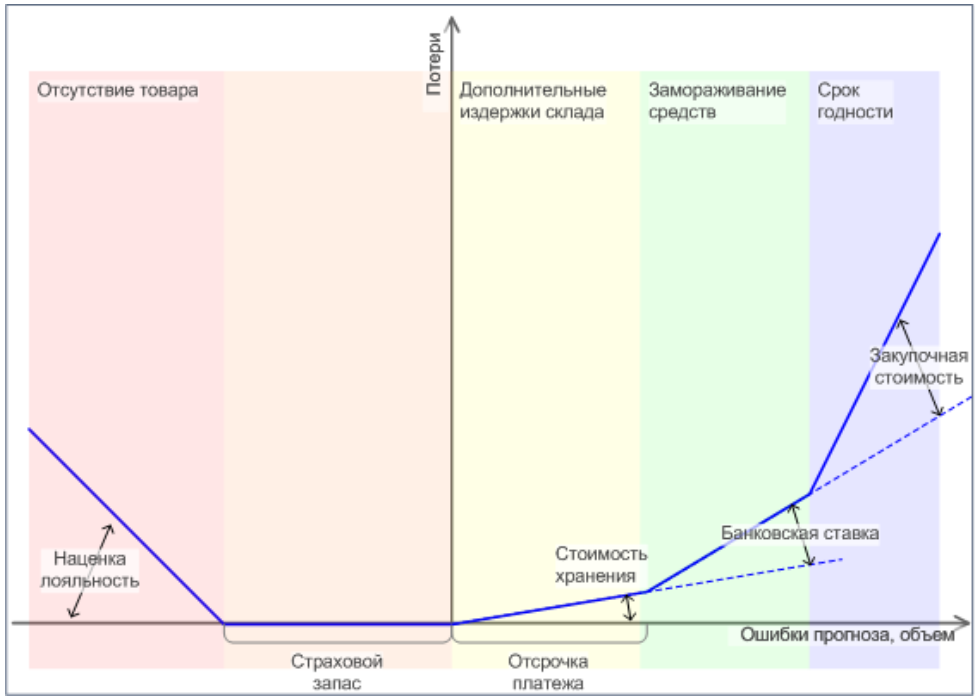
Объект — тройка $\langle$товар, магазин, день$\rangle$.

Примеры признаков:

- **бинарные:** выходной день, праздник, промоакция, и т. д.
- **количественные:** объёмы продаж в предшествующие дни.

Особенности задачи:

- функция потерь не квадратична и даже не симметрична;
- разреженные данные.



Конкурс kaggle.com: TFI Restaurant Revenue Prediction

Объект — место для открытия нового ресторана.

Предсказать — прибыль от ресторана через год.

Примеры признаков:

- демографические данные: возраст, достаток и т.д.,
- цены на недвижимость поблизости,
- маркетинговые данные: наличие школ, офисов и т.д.

Особенности задачи:

- мало объектов, много признаков;
- разнотипные признаки;
- есть выбросы;
- разнородные объекты (возможно, имеет смысл строить разные модели для мелких и крупных городов).

Машинное обучение на данных сложной структуры

- **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
- **Управление беспилотным аппаратом:**
  объект — поток данных с видеокамер, датчиков
  ответ — поток решений и управляющих сигналов

Предпосылки прорыва ИИ в задачах со сложными данными:

- большие и *чистые* данные (Big Data)
- методы оптимизации для задач большой размерности и обучаемая векторизация данных (Deep Learning)
- рост вычислительных мощностей (закон Мура, GPU)

Задача ранжирования поисковой выдачи

Объект — пара $\langle$текстовый запрос, документ$\rangle$.

Классы — релевантен или не релевантен,
разметка делается людьми — асессорами.

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.

Особенности данных и постановок прикладных задач

- разнородные (признаки измерены в разных шкалах)
- неполные (измерены не все, имеются пропуски)
- неточные (измерены с погрешностями)
- противоречивые (объекты одинаковые, ответы разные)
- избыточные (сверхбольшие, не помещаются в память)
- недостаточные (объектов меньше, чем признаков)
- сложно структурированные (нет признаковых описаний)

Риски, связанные с постановкой задачи:

- «грязные» данные
  (заказчик не обеспечивает качество данных)
- неясные критерии качества модели
  (заказчик не определился с целями или критериями)

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 $\{x_i : i = 1, \ldots, \ell\}$

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

Фрэнсис Бэкон. Новый органон. 1620.

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* («бритва 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

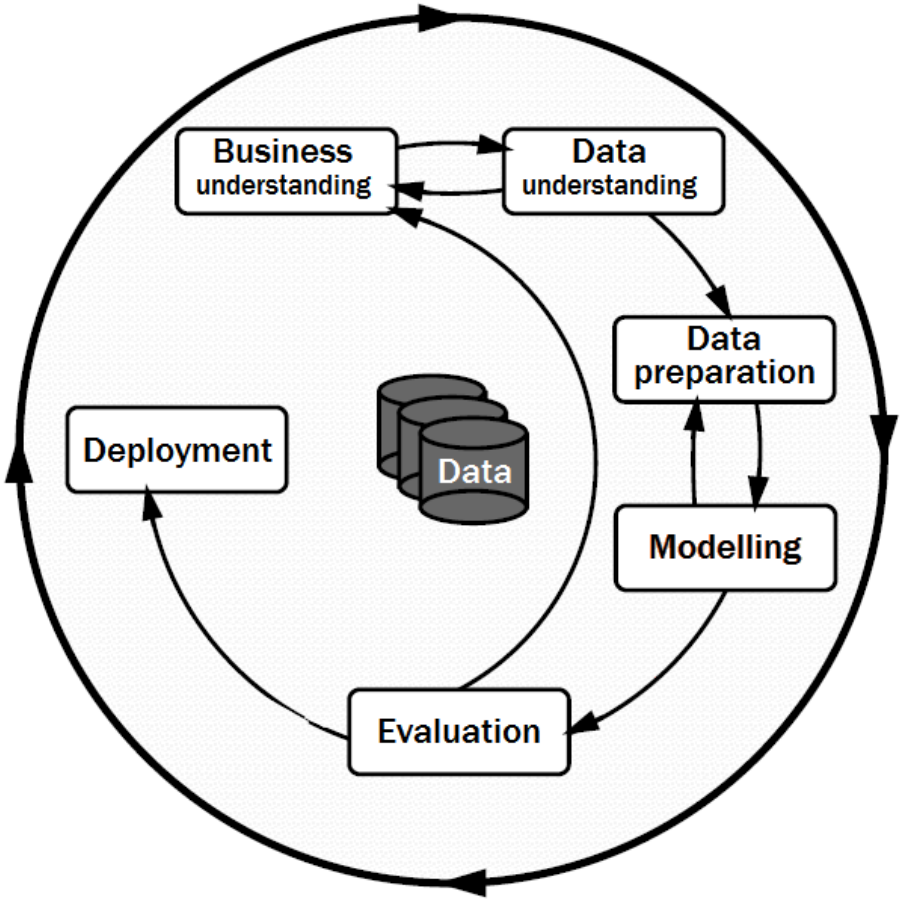
Наблюдения, измерения → выборка данных

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)

Межотраслевой стандарт интеллектуального анализа данных

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)



Компании-инициаторы:

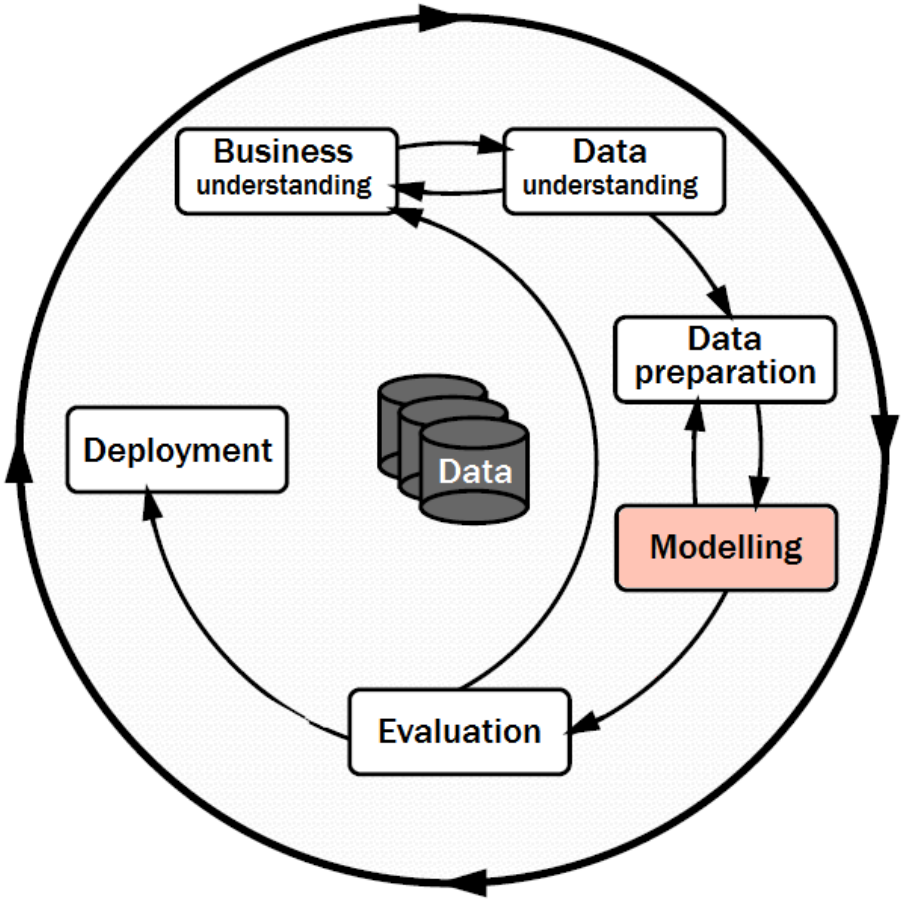
- SPSS
- Teradata
- Daimler AG
- NCR Corp.
- OHRA

Шаги процесса:

- понимание бизнеса
- понимание данных
- предобработка данных и инженерия признаков
- разработка моделей и настройка параметров
- оценивание качества
- внедрение

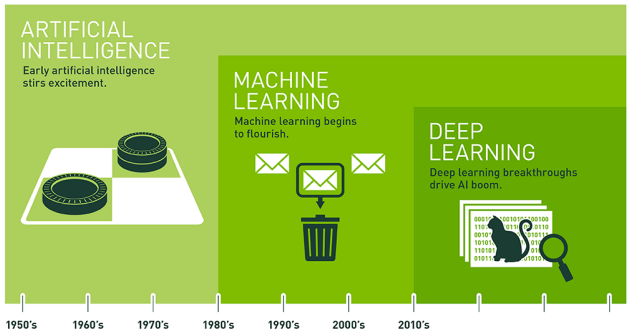
Понимание эволюции ИИ как автоматизации шагов CRISP-DM

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)



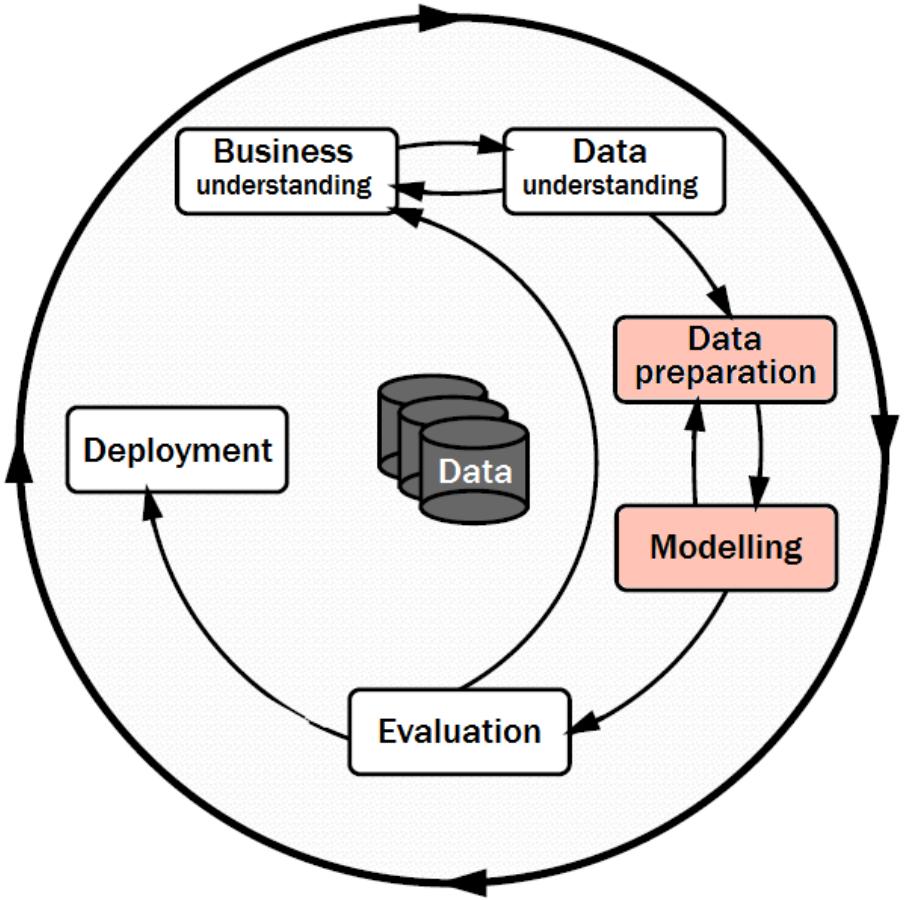
Эволюция ИИ:

- *Expert Systems:* жёсткие модели, основанные на правилах
- *Machine Learning:* параметрические модели, обучаемые по данным



Понимание эволюции ИИ как автоматизации шагов CRISP-DM

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)

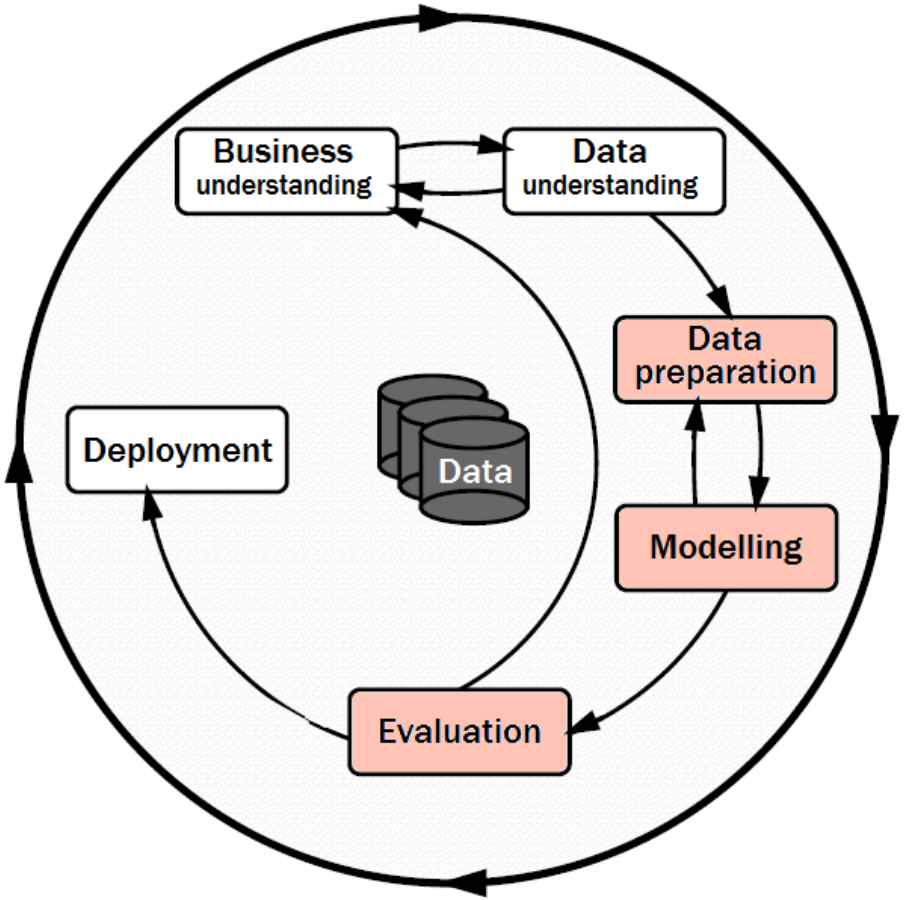


Эволюция ИИ:

- *Expert Systems:* жёсткие модели, основанные на правилах
- *Machine Learning:* параметрические модели, обучаемые по данным
- *Deep Learning:* модели с обучаемой векторизацией данных

Понимание эволюции ИИ как автоматизации шагов CRISP-DM

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)

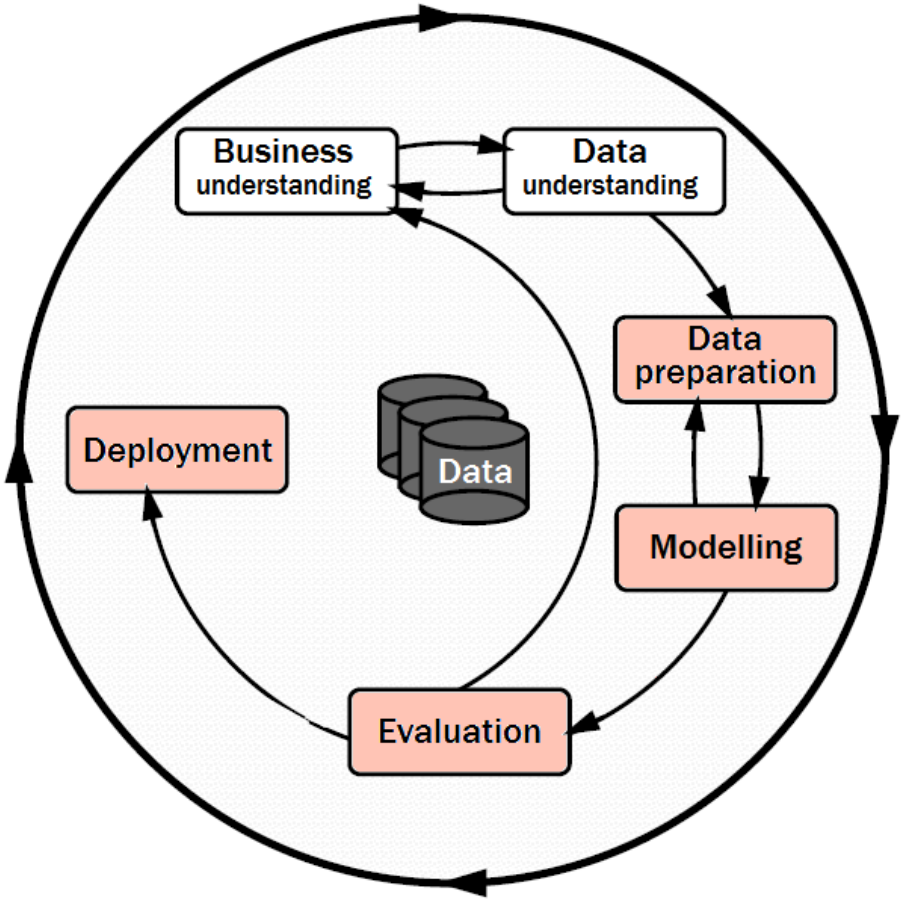


Эволюция ИИ:

- *Expert Systems:* жёсткие модели, основанные на правилах
- *Machine Learning:* параметрические модели, обучаемые по данным
- *Deep Learning:* модели с обучаемой векторизацией данных
- *AutoML:* автоматический выбор моделей и архитектур

Понимание эволюции ИИ как автоматизации шагов CRISP-DM

CRISP-DM: CRoss Industry Standard Process for Data Mining (1999)



Эволюция ИИ:

- *Expert Systems:* жёсткие модели, основанные на правилах
- *Machine Learning:* параметрические модели, обучаемые по данным
- *Deep Learning:* модели с обучаемой векторизацией данных
- *AutoML:* автоматический выбор моделей и архитектур
- *Lifelong Learning:* бесшовная интеграция обучения и выбора моделей в бизнес-процесс

Эксперименты на реальных данных

Эксперименты на конкретной прикладной задаче:

- цель — решить задачу как можно лучше
- важно понимание задачи и данных
- важно придумывать информативные признаки
- конкурсы по анализу данных: http://www.kaggle.com

Эксперименты на наборах прикладных задач:

- цель — протестировать метод в разнообразных условиях
- нет необходимости (и времени) разбираться в сути задач :(
- признаки, как правило, уже кем-то придуманы
- репозиторий UC Irvine Machine Learning Repository http://archive.ics.uci.edu/ml (682 задачи, 2025-09-01)

Эксперименты на синтетических данных

Используются для тестирования новых методов обучения.
Преимущество — мы знаем истинную $y(x)$ (ground truth)

Эксперименты на синтетических данных:

- цель — отладить метод, выявить границы применимости
- объекты x_i из придуманного распределения (часто 2D)
- ответы $y_i = y(x_i)$ для придуманной функции $y(x)$
- двумерные данные + визуализация выборки

Эксперименты на полу-синтетических данных:

- цель — протестировать помехоустойчивость модели
- объекты x_i из реальной задачи (признаки + шум)
- ответы $y_i = y(x_i)$ для придуманной функции $y(x)$ (+ шум)

Резюме в конце лекции

- **Основные понятия машинного обучения:**
  объект, ответ, признак, функция потерь, модель,
  метод обучения, эмпирический риск, переобучение

- **Постановка задачи** — это ДНК (Дано, Найти, Критерий)

- **Этапы решения задач машинного обучения:**
  — понимание задачи и данных
  — предобработка данных и изобретение признаков
  — конструирование параметрической модели
  — сведение обучения к оптимизации параметров
  — решение проблем оптимизации и переобучения
  — тестирование, оценивание качества
  — внедрение и эксплуатация

- **Прикладные задачи машинного обучения:**
  очень много, очень разных,
  во всех областях бизнеса, науки, производства