

Юрий Иванович Журавлёв. У истоков искусственного интеллекта

Воронцов Константин Вячеславович
k.vorontsov@iai.msu.ru

д.ф.-м.н., профессор РАН,
зав. каф. математических методов прогнозирования ВМК МГУ,
зав. лаб. машинного обучения и семантического анализа
Института искусственного интеллекта МГУ,
зав. каф. интеллектуальных систем МФТИ,
г.н.с. ФИЦ «Информатика и управление» РАН

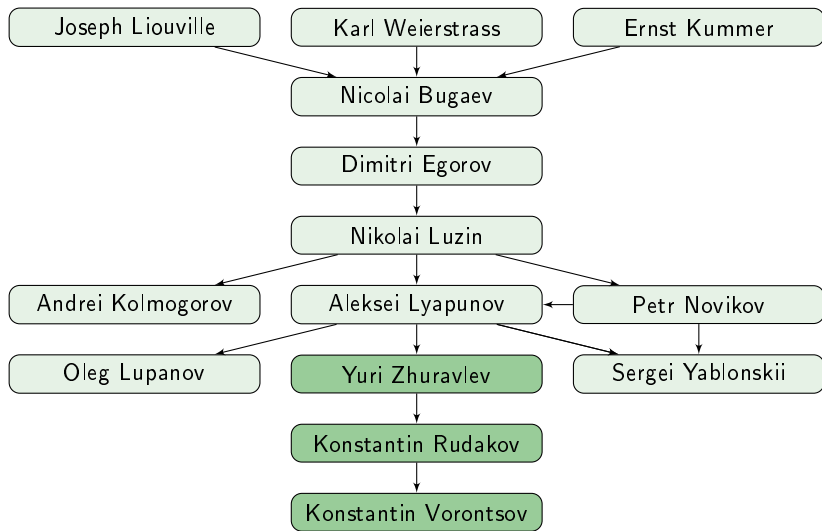
День математика • ВМК МГУ • 30 ноября 2025

- 1 ИИ: история развития и основные понятия**
 - задачи машинного обучения
 - докомпьютерная история ИИ
 - научные школы машинного обучения
- 2 Научная школа Юрия Ивановича Журавлёва**
 - алгоритмы вычисления оценок
 - задачи на малых данных
 - алгебраический подход к распознаванию
- 3 Настоящее и будущее искусственного интеллекта**
 - глубокое обучение как векторизация данных
 - большие языковые модели
 - явление эмерджентности

Юрий Иванович Журавлёв (1935–2022)

- 1957** окончил мехмат МГУ
- 1959** кандидат наук
- 1965** доктор наук (теория локальных алгоритмов)
- 1966** Ленинская премия в области науки и техники
- 1966** цикл работ по прогнозированию месторождений золота
- 1969** начало работы в ВЦ АН СССР
- 1971** алгоритмы вычисления оценок (АВО)
- 1977** алгебраический подход к проблеме распознавания образов
- 1984** член-корреспондент РАН
- 1986** премия Совета министров СССР
- 1992** академик РАН
- 1997** организовал кафедру ММП ВМК МГУ





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.

Задача машинного обучения (supervised machine learning)

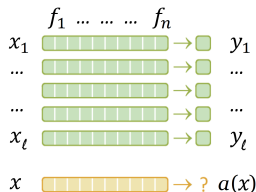
Дано: обучающая выборка

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 модели a при правильном ответе y .

Основные типы задач машинного обучения с учителем:

- $\mathcal{L}(a, y) = (a - y)^2$ в задачах регрессии, $y_i \in \mathbb{R}$
- $\mathcal{L}(a, y) = [a \neq y]$ в задачах классификации, $y_i \in \{0, 1\}$

Метод наименьших квадратов (Гаусс, 1795)

Линейная модель регрессии:

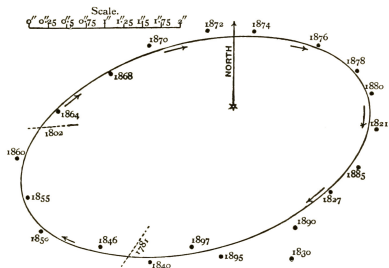
$$a(x, w) = \sum_{j=1}^n w_j f_j(x), \quad w \in \mathbb{R}^n$$

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.

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

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

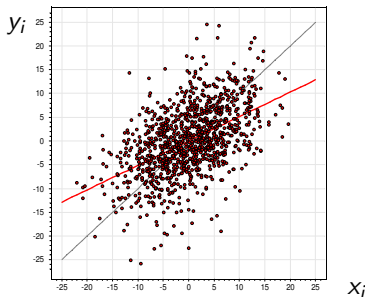
Критерий

наименьших квадратов:

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

Решение:

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

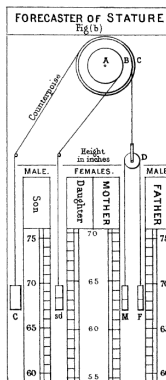
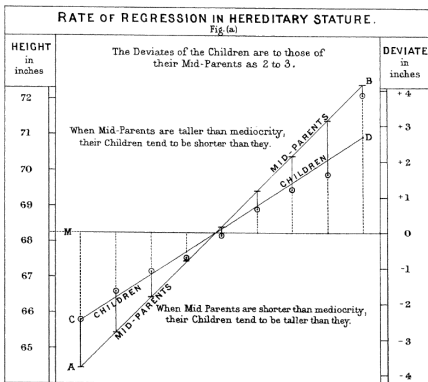


Фрэнсис
Гальтон
(1822–1911)

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

История термина «регрессия» (Гальтон, 1886)

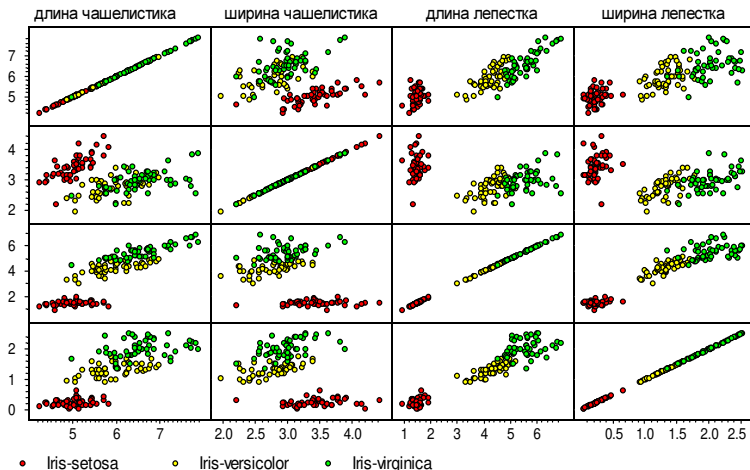
«Регрессия к посредственности» — угол наклона меньше 1
 Скрытый смысл: обратный ход исследования от данных к модели



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

Задача классификации цветков ириса (Фишер, 1936)

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



Линейный дискриминантный анализ (Фишер, 1936)

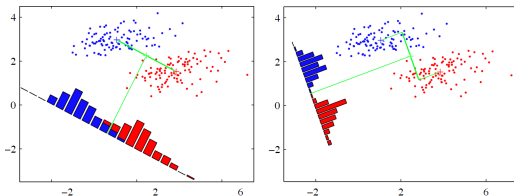
Найти линейную модель классификации:

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

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



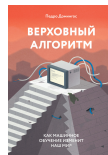
Рональд
Фишер
(1890–1962)



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

Основные научные школы машинного обучения

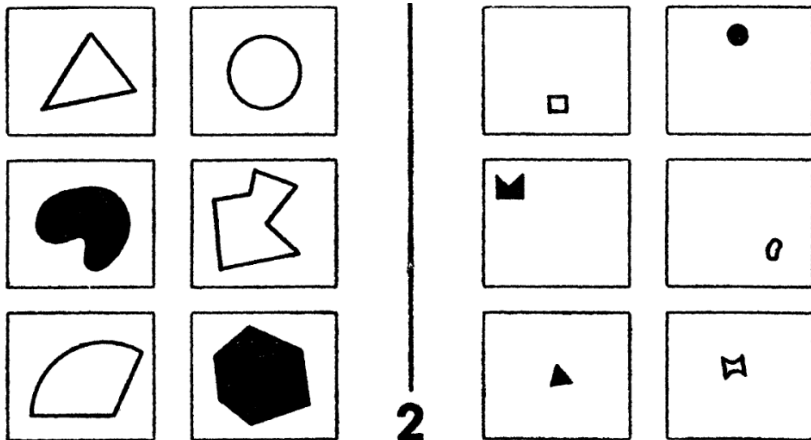
- 1 *символизм* – поиск логических закономерностей в данных
 - Decision Tree, Rule Induction
- 2 *коннекционизм* – обучение искусственных нейронных сетей
 - BackPropagation, Deep Belief Nets, Deep Learning
- 3 *аналогизм* – «близкие объекты имеют близкие ответы»
 - kNN, RBF, SVM, Kernel Smoothing
- 4 *эволюционизм* – самоорганизация сложных моделей
 - Genetic Algorithms, Genetic Programming, Symbolic Regression
- 5 *байесионизм* – оценивание распределений параметров
 - Naive Bayes, Bayesian Networks, Graphical Models
- ⊕ *композиционизм* – ансамбли моделей
 - Weighted Voting, Boosting, Bagging, Stacking, Random Forest, Яндекс.CatBoost



Педро Домингос. Верховный алгоритм. 2016. 336 с.

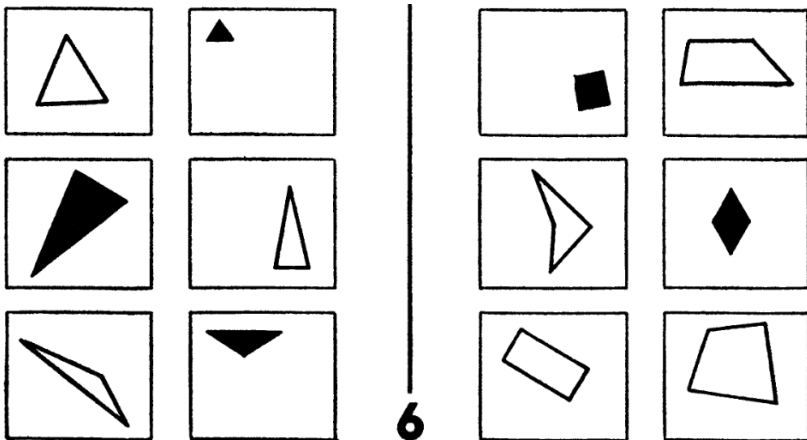
Тесты Бонгарда [Проблема узнавания, 1967]

Обучающая выборка: по 6 объектов каждого из двух классов.
Требуется найти правило классификации.



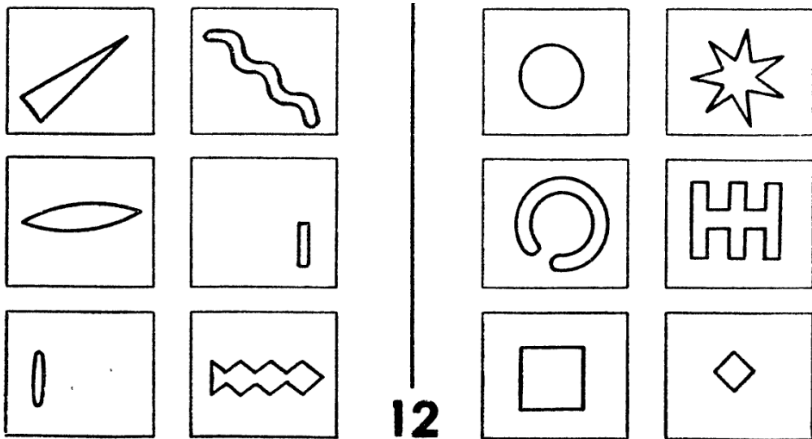
Тесты Бонгарда [Проблема узнавания, 1967]

Что даёт нам уверенность, что мы нашли верное правило?
Безошибочная классификация примеров обучающей выборки.



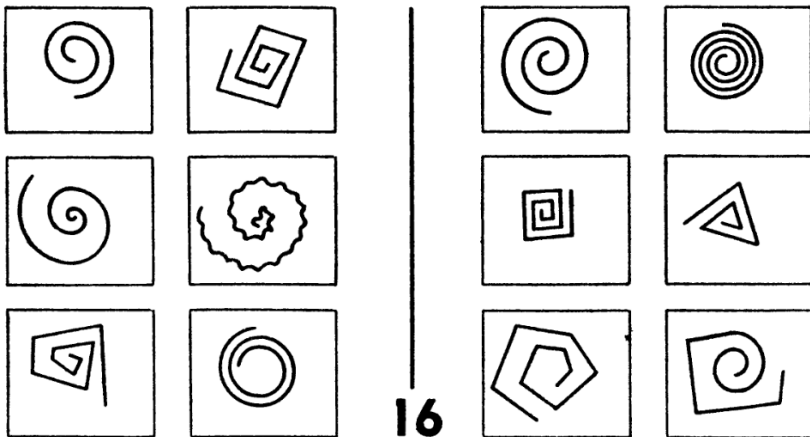
Тесты Бонгарда [Проблема узнавания, 1967]

Что ещё даёт нам уверенность, что мы нашли верное правило?
Простота, общность, лаконичность найденного правила.



Тесты Бонгарда [Проблема узнавания, 1967]

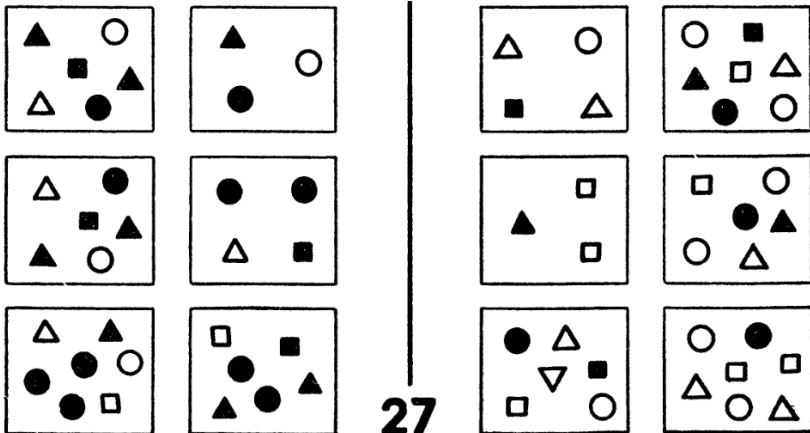
Мы решаем эти задачи почти мгновенно. Чем мы пользуемся?
Однако для компьютера они сложны. Чего ему не хватает?



Тесты Бонгарда [Проблема узнавания, 1967]

Нужно ли закладывать знания геометрии в явном виде?

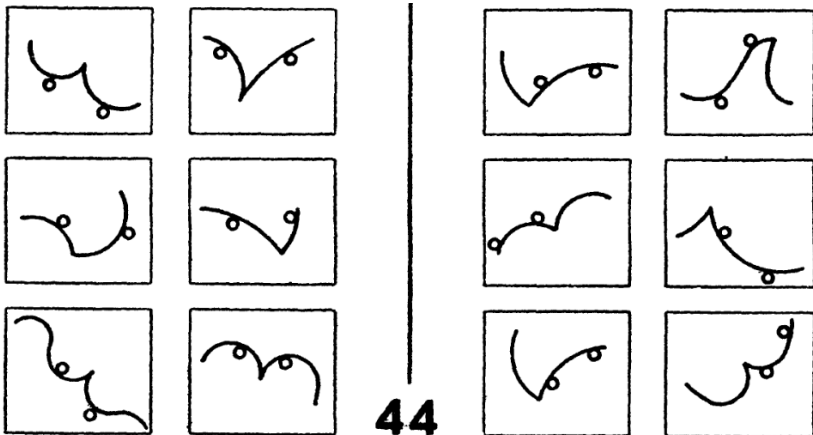
Или возможно выработать необходимые понятия на примерах?



Тесты Бонгарда [Проблема узнавания, 1967]

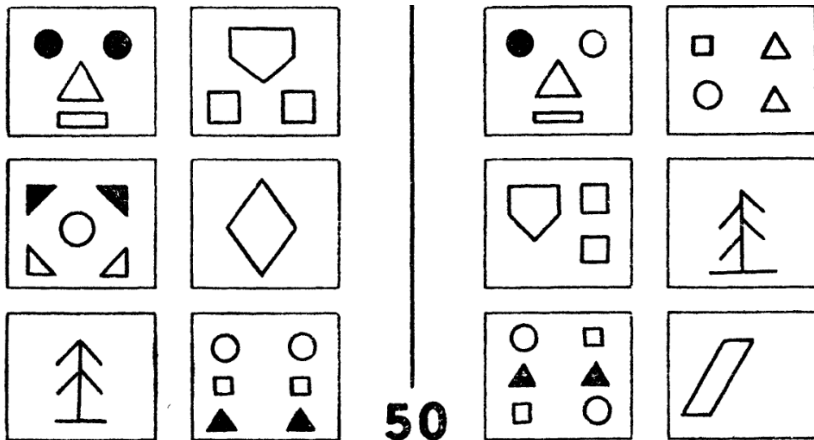
Как вычислять полезные признаки по «сырым» данным?

Возможно ли поручить перебор признаков и моделей машине?



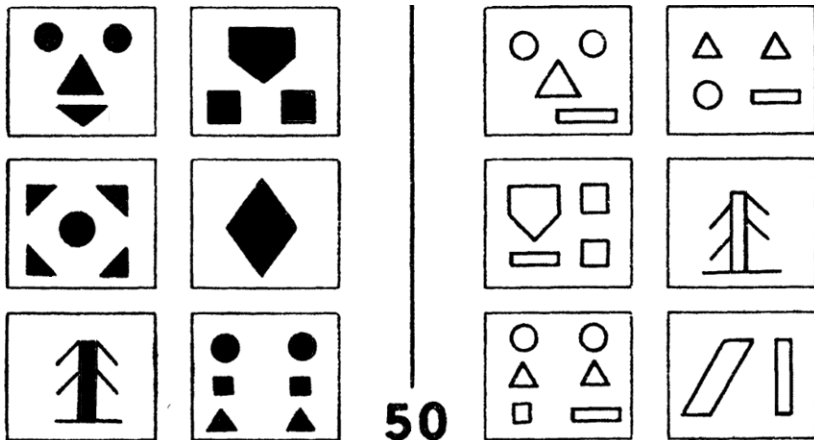
Тесты Бонгарда [Проблема узнавания, 1967]

Каков риск вывести из данных ложное правило, *предрассудок*?
Как этот риск зависит от числа примеров и сложности правил?



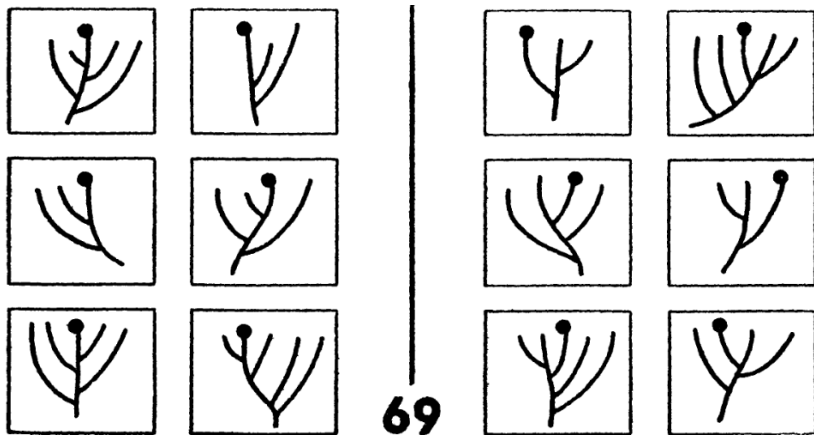
Тесты Бонгарда [Проблема узнавания, 1967]

Какого числа примеров достаточно для выработки правила?
Что делать, если к выборке подходит много разных правил?



Тесты Бонгарда [Проблема узнавания, 1967]

Эти вопросы составляют основу машинного обучения сегодня.
М.М.Бонгард поставил все эти проблемы в середине 60-х!



Символизм. Научная школа М. М. Бонгарда

- 1958: Программа «Открой закон» восстанавливала зависимость полным комбинаторным перебором формул
- 1959: Программа «Арифметика» для сокращения перебора использовала оценки *информативности*
- 1961: Программа «КоРа» перебирала *информативные тройки* признаков



Михаил Моисеевич
Бонгард
(1924–1971)

«КоРа-3»: первое применение распознавания незрительных образов для распознавания границы нефть-вода в скважине. Введены принципы *голосования*, *скользящего контроля*, понятия *информативности* и *предрассудка* (переобучения).

Бонгард М. М., Вайнцвайг М. Н., Губерман Ш. А. Извекова М. Л., Смирнов М. С. Использование обучающейся программы для выявления нефтеносных пластов. 1966.

Понятие информативной логической закономерности

Модель классификации — взвешенное голосование правил:

$$a(x, w) = \arg \max_{y \in Y} \sum_{k=1}^{n_y} w_{yk} R_{yk}(x)$$

Правило — конъюнкция элементарных пороговых условий:

$$R(x) = \bigwedge_{j \in \omega} [a_j \leq f_j(x) \leq b_j]$$

Синдром — выполнены не менее d условий из множества ω ,

$$R(x) = \left[\sum_{j \in \omega} [a_j \leq f_j(x) \leq b_j] \geq d \right]$$

Правило R является закономерностью класса $y \in Y$, если

$$\begin{cases} p_y(R) = \#\{x_i: R(x_i)=1 \text{ и } y_i=y\} \rightarrow \max \\ n_y(R) = \#\{x_i: R(x_i)=1 \text{ и } y_i \neq y\} \rightarrow \min \end{cases}$$

Требование интерпретируемости

- 1) $R(x)$ записывается на естественном языке
- 2) $R(x)$ зависит от небольшого числа признаков (не более 7)

Пример (из области медицины)

Если «возраст > 60 » и «пациент ранее перенёс инфаркт»,
то операцию не делать, риск отрицательного исхода 60%

Пример (из области кредитного скоринга)

Если «в анкете указан домашний телефон»
и «зарплата $> \$2000$ » и «сумма кредита $< \$5000$ »
то кредит можно выдать, риск дефолта 5%

Замечание. *Риск* — частотная оценка вероятности класса, вычисляемая, как правило, по отложенной контрольной выборке

Коннекционизм. Линейная модель нейрона

Линейная модель нейрона (1943):

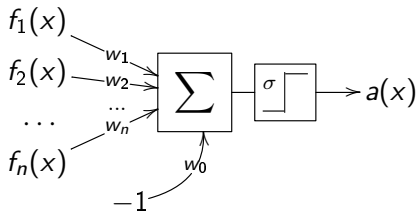
$$a(x, w) = \sigma \left(\sum_{j=1}^n w_j f_j(x) - w_0 \right)$$

$f_j(x)$ — признаки объекта x

w_j — веса признаков

w_0 — порог активации

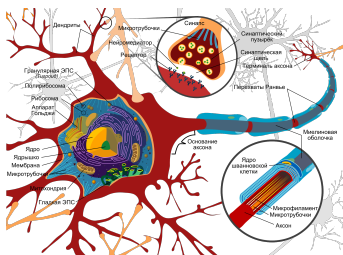
$\sigma(z)$ — функция активации



Уоррен
МакКаллок
(1898–1969)



Вальтер
Питтс
(1923–1969)



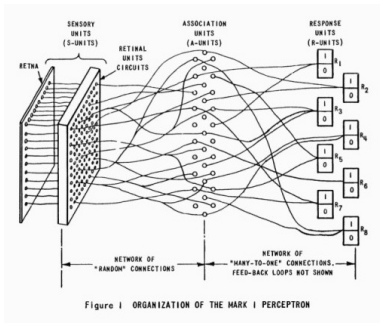
Персептрон Розенблатта (1957)

Mark-1 — первый нейрокомпьютер (1960)

для распознавания цифр и фигур

Обучение — метод коррекции ошибки

Архитектура — двухслойная сеть



Фрэнк Розенблатт
(1928–1971)



Розенблатт Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. 1965 (1962)

Двухслойные сети — универсальные аппроксиматоры функций

Функция $\sigma(z)$ — сигмоида, если $\lim_{z \rightarrow -\infty} \sigma(z) = 0$ и $\lim_{z \rightarrow +\infty} \sigma(z) = 1$.

Теорема Цыбенко (universal approximation theorem, 1989)

Если $\sigma(z)$ — непрерывная сигмоида, то для любой непрерывной на $[0, 1]^n$ функции $f(x)$ существуют такие значения параметров H , $\alpha_h \in \mathbb{R}$, $w_h \in \mathbb{R}^n$, $w_0 \in \mathbb{R}$, что двухслойная сеть

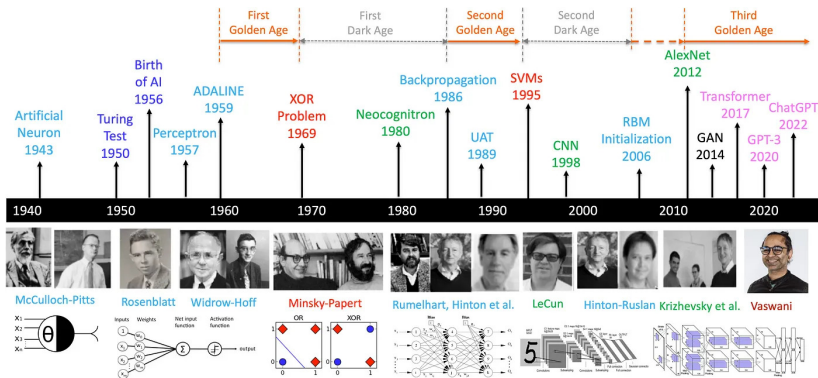
$$a(x) = \sum_{h=1}^H \alpha_h \sigma(\langle x, w_h \rangle - w_0)$$

равномерно приближает $f(x)$ с любой точностью ε :

$$|a(x) - f(x)| < \varepsilon, \text{ для всех } x \in [0, 1]^n.$$

George Cybenko. Approximation by Superpositions of a Sigmoidal function. Mathematics of Control, Signals, and Systems. 1989.

Основные вехи развития нейронных сетей (AI winters)



Минский М., Пайперт С. Перцептроны. 1971 (1969)

Галушкин А. И. Синтез многослойных систем распознавания образов. 1974

Ива́хненко А. Г., Лапа В. Г. Кибернетические предсказывающие устройства. 1965

Rummelhart D. et al. Learning internal representations by error propagation. 1986

Krizhevsky A. et al. ImageNet classification with deep convolutional neural networks. 2012

Vaswani A. et al. Attention is all you need. 2017

Аналогизм. Научная школа М. А. Айзермана

- *Гипотеза компактности*: схожие объекты, как правило, находятся в одном классе
- *Метод потенциальных функций*: идея заимствуется из физики
- *Линейная модель классификации*: взвешенное голосование функций сходства $f_i(x) = K(x, x_i)$ между x и x_i :

$$a(x) = \arg \max_{y \in Y} \sum_{i: y_i=y} \alpha_{yi} K(x, x_i)$$



Марк Аронович
Айзерман
(1913–1992)

Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Теоретические основы метода потенциальных функций в задаче об обучении автоматов разделению входных ситуаций на классы. 1964.

Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. 1970.

Аркадьев А. Г., Браверман Э. М. Обучение машин распознаванию образов. 1964.

Научная школа В. Н. Вапника и А. Я. Червоненкиса

Семейство классификаторов A обучаемо:

$$P\left\{\sup_{a \in A} |P(a) - \nu(a, X^\ell)| > \varepsilon\right\} \leq \eta,$$

$P(a)$ — вероятность ошибки классификатора,
 $\nu(a, X^\ell)$ — эмпирический риск — частота
 ошибок классификатора a на выборке.

Основные результаты VC-теории:

- Обосновано ограничение сложности A
- Понятие ёмкости семейства, $VCdim$
- Метод структурной минимизации риска
- Метод опорных векторов, SVM



Владимир
Наумович Вапник

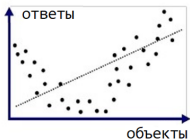


Алексей Яковлевич
Червоненкис
(1938–2014)

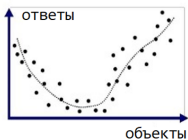
Вапник В. Н., Червоненкис А. Я.

Теория распознавания образов. М.: Наука, 1974.

Проблемы недообучения и переобучения

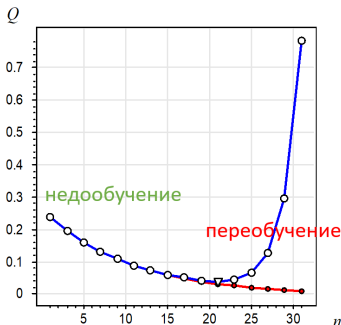


недообучение



переобучение

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



Эволюционизм. Научная школа А. Г. Ива́хненко

Метод группового учёта аргументов (МГУА)

основан на *самоорганизации моделей*

— переборной оптимизации структуры модели

- задача *отбора признаков* (feature selection)
- качество моделей оценивается в процессе перебора по многим *внешним критериям*:
 - скользящий контроль
 - помехоустойчивость моделирования
 - баланс / согласованность прогнозов и др.
- первая 8-слойная глубокая нейросеть (1965)
- сотни применений, около 300 диссертаций



Алексей
Григорьевич
Ива́хненко
(1913–2007)

Ивахненко А. Г., Лапа В. Г. Кибернетические предсказывающие устройства. 1965.

Ивахненко А. Г., Зайченко Ю. П., Димитров В. Д. Принятие решений на основе самоорганизации. 1976.

Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем. 1982.

Алгоритмы вычисления оценок, АВО (Ю. И. Журавлёв)

Объединение основных на тот момент нестрогих (эвристических) принципов:

- символизм (вывод правил из данных)
- эволюционизм (отбор наилучших правил)
- аналогизм (оценки сходства объектов)
- байесионизм (классы — смеси плотностей)
- коннекционизм (взвешенное голосование)

$$a(x) = \arg \max_{y \in Y} \sum_{i: y_i=y} \sum_{\omega \in \Omega} w_{\omega i} B_{\omega i}(x, x_i)$$

где $B_{\omega i}$ — бинарные функции сходства по наборам признаков ω :

$$B_{\omega i}(x, x_i) = \bigwedge_{j \in \omega} [|f_j(x) - f_j(x_i)| < \varepsilon]$$



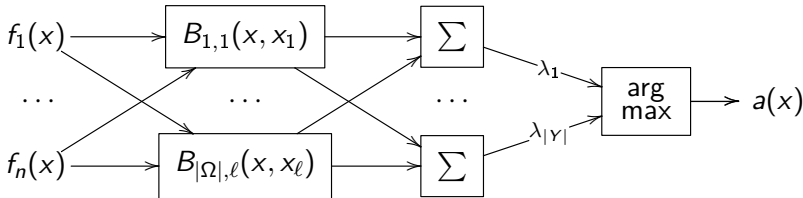
Юрий
Иванович
Журавлёв
(1935–2022)

Журавлёв Ю. И., Никифоров В. В. Алгоритмы распознавания, основанные на вычислении оценок, 1971.

АВО объединяет многие эвристические принципы

≈ трёхслойная нейросеть RBF (Radial Basis Function):

$$a(x) = \arg \max_{y \in Y} \lambda_y \sum_{i, \omega} [y_i = y] w_{\omega i} B_{\omega i}(x, x_i)$$



≈ метод потенциальных функций $K(x, x_i) = B_{\omega i}(x, x_i)$

≈ линейный классификатор SVM с радиальным ядром

≈ байесовский классификатор с плотностями-смесями $p(x|y)$

≈ отбор эталонов: $w_{\omega i} = 0$ для не-эталонов x_i

≈ отбор признаков в бинарных функциях сходства

Принципы информативности, непротиворечивости, тупиковости

- *информативность* предиката $R(x)$ класса $y \in Y$:

$$\begin{cases} p_y(R) = \#\{x_i: R(x_i)=1 \text{ и } y_i=y\} \rightarrow \max \\ n_y(R) = \#\{x_i: R(x_i)=1 \text{ и } y_i \neq y\} \rightarrow \min \end{cases}$$
- *информативность* функции сходства $B(x, x')$:

$$\begin{cases} p(B) = \#\{(x_i, x_j): B(x_i, x_j)=1 \text{ и } y_i=y_j\} \rightarrow \max \\ n(B) = \#\{(x_i, x_j): B(x_i, x_j)=1 \text{ и } y_i \neq y_j\} \rightarrow \min \end{cases}$$
- *непротиворечивость*: $n_y(b) = 0, n(B) = 0$
 - *тест* ω : $B_\omega(x_i, x_j) = 0, \forall i, j: y_i \neq y_j$
 - *представительный набор* (ω, i) : $B_\omega(x_i, x_j) = 0, \forall j: y_i \neq y_j$
- *тупиковость*: никакое подмножество признаков $\omega' \subset \omega$ не является тестом (или представительным набором)

Дмитриев А. Н., Журавлев Ю. И., Кренделев Ф. П. Об одном принципе классификации и прогноза геологических объектов и явлений. 1968.

Журавлёв Ю. И., Никифоров В. В. Алгоритмы распознавания, основанные на вычислении оценок, 1971.

Задачи на малых данных

Особенности геологических данных в задачах поиска месторождений редкого типа (золото, уран, алмазы и т.д.)

- объектов мало (7 + 11), признаков много (более сотни)
- надёжной геофизической модели не существует
- в данных бывают «пропуски» — неизмеренные значения

Группа признаков		Пространственно-временные						Вещественные						
		2	3	4	6	69	21	22	23	24
Месторождения M_i	Витватерсранд (M^1_i)	1	0	0	1	0	1	1	1	0	1	0	0	1
	Блайнд-Ривер (M^2_i)	1	0	0	1	0	0	0	1	1	1	0	0	1
	Жакобина (M^3_i)	0	0	1	0	1	0	1	1	1	1	0	1	1
	Муанга, Габон (M^4_i)	0	1	0	0	1	0	—	1	1	0	0	1	0
	Тарква, Гапа (M^5_i)	1	0	0	0	0	0	—	1	1	1	0	0	1
	Австралия (M^6_i)	0	0	1	—	0	1	—	1	—	—	—	1	1
	Эпо-Кюлия, Финляндия (M^7_i)	1	0	0	1	1	1	1	—	—	—	1	—	1

Кренделев Ф. П., Дмитриев А. Н., Журавлев Ю. И. Сравнение геологического строения зарубежных месторождений докембрийских конгломератов с помощью дискретной математики. Доклады АН СССР. 1967

Дмитриев А. Н., Журавлев Ю. И., Кренделев Ф. П. Об одном принципе классификации и прогноза геологических объектов и явлений. 1968.

Ансамблирование предсказательных моделей

$X^\ell = (x_i, y_i)_{i=1}^\ell \subset X \times Y$ — обучающая выборка, $y_i = y(x_i)$

$a_t: X \rightarrow Y$, $t = 1, \dots, T$ — обучаемые базовые алгоритмы

Идея ансамблирования (Ю.И.Журавлёв): как из множества по отдельности плохих алгоритмов a_t построить один хороший?

Декомпозиция базовых алгоритмов $a_t(x) = C(b_t(x))$

$a_t: X \xrightarrow{b_t} R \xrightarrow{C} Y$, где R — удобное пространство оценок,

b_t — базовые алгоритмические операторы,

C — решающее правило простого вида.

Ансамбль (композиция) базовых алгоритмов a_1, \dots, a_T ,

$F: R^T \rightarrow R$ — корректирующая (агрегирующая) операция

$$a(x) = C(F(b_1(x), \dots, b_T(x)))$$

Ю.И.Журавлёв. Об алгебраическом подходе к решению задач распознавания или классификации. Проблемы кибернетики, 1978.

Агрегирующие (корректирующие) функции

Общие требования к агрегирующей функции:

- $F(b_1, \dots, b_T, x) \in [\min_t b_t, \max_t b_t]$ — среднее по Коши $\forall x$
- $F(b_1, \dots, b_T, x)$ монотонно не убывает по всем b_t

Примеры агрегирующих функций:

- простое голосование (simple voting):

$$F(b_1, \dots, b_T) = \frac{1}{T} \sum_{t=1}^T b_t$$

- взвешенное голосование (weighted voting):

$$F(b_1, \dots, b_T) = \sum_{t=1}^T \alpha_t b_t, \quad \sum_{t=1}^T \alpha_t = 1, \quad \alpha_t \geq 0$$

- смесь алгоритмов (mixture of experts)

с функциями компетентности (gating function) $g_t: X \rightarrow \mathbb{R}$

$$F(b_1, \dots, b_T, x) = \sum_{t=1}^T g_t(x) b_t(x)$$

Обучение предсказательных моделей и их ансамблей

$\mathcal{L}(b, x_i)$ — функция потерь модели $b(x_i, w)$ при ответе y_i

Минимизация эмпирического риска для базовых алгоритмов:

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

Минимизация эмпирического риска для добавления базового алгоритма b_T в ансамбль при фиксации предыдущих:

$$\sum_{i=1}^{\ell} \mathcal{L}\left(\sum_{t=1}^{T-1} \alpha_t b_t(x_i, w_t) + \alpha_T b_T(x_i, w_T), y_i\right) \rightarrow \min_{\alpha_T, w_T}$$

Ю.И.Журавлёв. Корректные алгебры над множествами некорректных (эвристических) алгоритмов (I, II, III). Кибернетика, Киев, 1977–1978.

M.Kearns, L.G.Valiant. Cryptographic limitations on learning Boolean formulae and finite automata. 1989.

Y.Freund, R.E.Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. 1995.

К.В.Рудаков, К.В.Воронцов. О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания. Доклады РАН, 1999.

Композиции обучаемых моделей

- **Простое и взвешенное голосование**
Мазуров В. Д. Комитеты системы неравенств и задача распознавания. 1971.
Журавлёв Ю. И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. 1977.
Freund Y., Schapire R. E. A decision-theoretic generalization of on-line learning and an application to boosting. 1995.
Friedman G. Greedy Function Approximation: A Gradient Boosting Machine. 1999.
- **Случайный лес**
Breiman L. Random Forests. 2001.
- **Восстановление смесей распределений, EM-алгоритм**
Шлезингер М. И. О самопроизвольном различении образов. 1965.
Dempster A. P., Laird N. M., Rubin D. B. Maximum likelihood from incomplete data via the EM-algorithm. 1977.
- **Смеси классификаторов с областями компетентности**
Растринин Л. А., Эренштейн Р. Х. Коллективные правила распознавания. 1981.
Jacobs R. A., Jordan M. I., Nowlan S. J., Hinton G. E. Adaptive mixtures of local experts. 1991.

Градиентный бустинг и случайный лес — универсальные и наиболее успешные методы классификации.

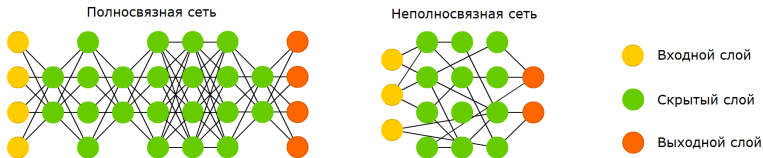
MatrixNet и *CatBoost* — эффективные реализации от Яндекса.

Глубокие нейронные сети (Deep Neural Network, DNN)

1965: первые глубокие нейронные сети

1997: рекуррентная сеть LSTM для анализа последовательностей

2012: свёрточная сеть для классификации изображений AlexNet



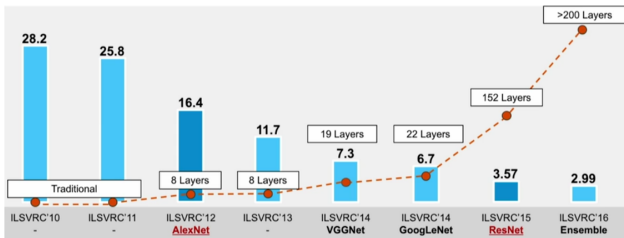
- *Архитектура сети* — структура слоёв и связей между ними, позволяющая наделять DNN нужными свойствами
- DNN позволяют принимать на входе и генерировать на выходе *сложно структурированные данные*

Ива́хненко А. Г., Лапа В. Г. Кибернетические предсказывающие устройства. 1965

Hochreiter S., Schmidhuber J. Neural Computation, 9(8), 1997

Krizhevsky A. et al. ImageNet classification with deep convolutional neural networks. 2012

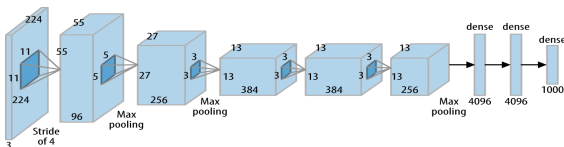
Глубокие свёрточные сети для классификации изображений



Старт в 2009

Человеческий уровень ошибок 5% пройден в 2015

Свёрточные
нейронные сети
AlexNet (2012)
ResNet (2015)



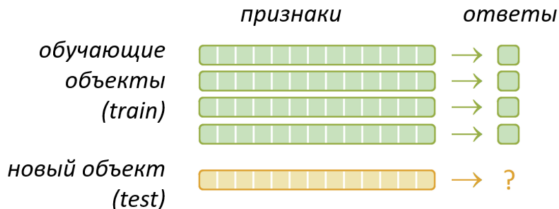
Li Fei-Fei et al. ImageNet: A large-scale hierarchical image database. 2009
Krizhevsky A. et al. ImageNet classification with deep convolutional neural networks. 2012
Kaiming He et al. Deep residual learning for image recognition. 2015

Три основных этапа. Этап 1: вектор → скаляр

Предсказательное моделирование векторных данных

Вход: векторные признаковые описания объектов

Выход: скалярные ответы (предсказания, прогнозы)



Приложения: медицинская диагностика,
геологическое прогнозирование, кредитный скоринг,...

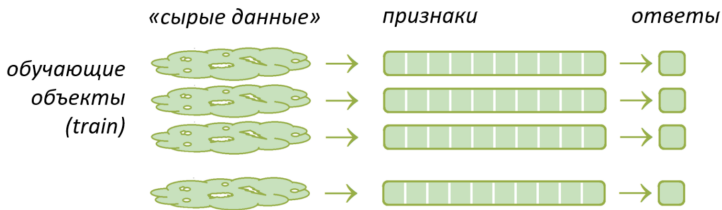
Модели: SVM, LR, MVR, RBF, MLP, ID3, CART, RF, GBM,...

Три основных этапа. Этап 2: структура → вектор → скаляр

Обучаемая векторизация сложно структурированных данных

Вход: сложно структурированные «сырые» данные объектов

Выход: векторные представления объектов, затем ответы



Приложения: классификация изображений, текстов, сигналов, голосовых команд, биометрическая идентификация личности,...

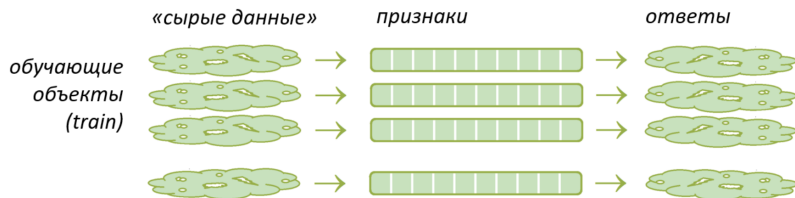
Модели: CNN, AlexNet, ResNet, word2vec, FastText, BERT,...

Три основных этапа. Этап 3: структура → вектор → структура

Обучаемая генерация сложно структурированных данных

Вход: сложно структурированные объекты

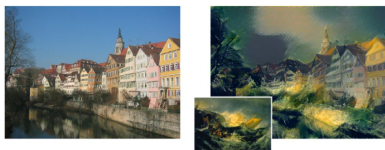
Выход: сложно структурированные ответы



Приложения: синтез изображений, перенос стиля, распознавание речи, машинный перевод, суммаризация текстов, чат-боты,...

Модели: seq2seq, RNN, LSTM, GAN, VAE, GPT,...

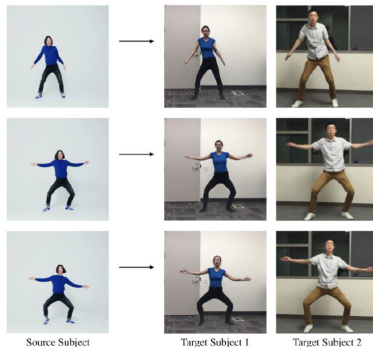
Примеры GAN для синтеза изображений и видео



(d) input image

(e) output 3d face

(f) textured 3d face



Source Subject

Target Subject 1

Target Subject 2

Chuan Li, Michael Wand. Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. 2016.

Xiaoxing Zeng, Xiaojiang Peng, Yu Qiao. DF2Net: A Dense Fine Finer Network for Detailed 3D Face Reconstruction. ICCV-2019.

Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros. Everybody Dance Now. ICCV-2019.

Эволюция подходов машинного обучения в анализе текстов

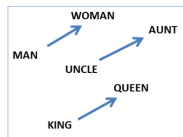
Декомпозиция задач по уровням пирамиды NLP

- морфологический анализ, лемматизация, опечатки
- синтаксический анализ, выделение терминов, NER
- семантический анализ, выделение фактов, тем



Модели векторных представлений (эмбедингов) слов на основе матричных разложений

- модели дистрибутивной семантики: word2vec [Mikolov, 2013], FastText [Bojanowski, 2016]
- тематические модели LDA [Blei, 2003], ARTM [2014]

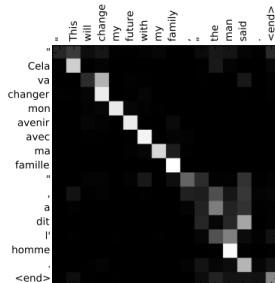
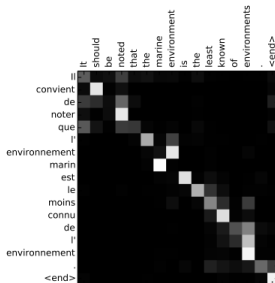
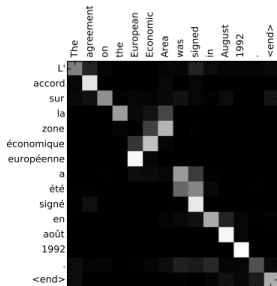


Нейросетевые модели локальных контекстов

- рекуррентные нейронные сети
- модели внимания и трансформеры: BERT [2018], GPT-3 [2020], GPT-4 [2023]

$$\text{softmax} \left(\frac{\begin{matrix} Q \\ \text{grid} \end{matrix} \times \begin{matrix} K^T \\ \text{grid} \end{matrix}}{\sqrt{d}} \right) \begin{matrix} V \\ \text{grid} \end{matrix}$$

Модели внимания для машинного перевода



Вход: $\{x_i\}$ — последовательность слов входного языка

Выход: $\{y_t\}$ — последовательность слов выходного языка

Интерпретация: матрица a_{it} показывает, на какие слова x_i модель обращает внимание, генерируя слово перевода y_t

Модели внимания для аннотирования изображений



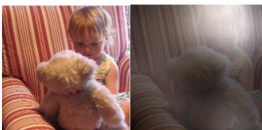
A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with trees in the background.

Подсвечены области, на которые модель обращает внимание, когда генерирует подчёркнутое слово в аннотации изображения

Kelvin Xu et al. Show, attend and tell: neural image caption generation with visual attention. 2016

Трасформер для машинного перевода

Трасформер (transformer) — это нейросетевая архитектура на основе моделей внимания и полносвязных слоёв

Схема преобразований данных в машинном переводе:

- $S = (w_1, \dots, w_n)$ — слова предложения на входном языке
↓ обучаемая или пред-обученная векторизация слов
- $X = (x_1, \dots, x_n)$ — векторы слов входного предложения
↓ трансформер-кодировщик
- $Z = (z_1, \dots, z_n)$ — контекстные векторы слов
↓ трансформер-декодировщик, похож на кодировщика
- $Y = (y_1, \dots, y_m)$ — векторы слов выходного предложения
↓ генерация слов из построенной языковой модели
- $\tilde{S} = (\tilde{w}_1, \dots, \tilde{w}_m)$ — слова предложения на выходном языке

Vaswani et al. (Google) Attention is all you need. 2017.

ChatGPT и GPT-4: проблески общего искусственного интеллекта

Sparks of Artificial General Intelligence: Early experiments with GPT-4

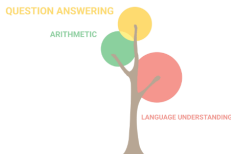
Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research (27 March 2023)

Новые способности модели, не закладывавшиеся при обучении:

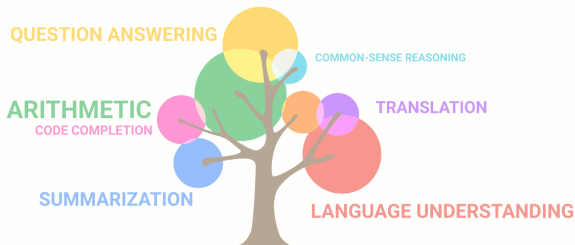
- объяснять свои ответы, перефразировать
- реферировать, генерировать планы, сценарии, шаблоны
- переводить на другие языки, строить аналогии, менять тональность, стиль, глубину изложения
- генерировать программный код на различных языках
- решать некоторые логические и математические задачи
- искать и исправлять собственные ошибки по подсказке

Появление у модели качественно новых способностей



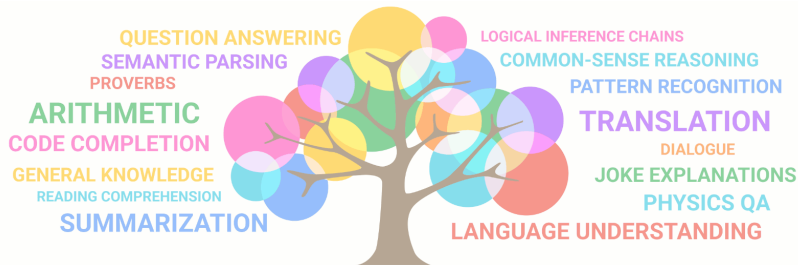
- GPT-2: 14/Feb/2019, контекст 768 слов (1,5 страницы)
- 1,5 млрд. параметров, корпус 10 млрд. токенов (40Gb)
- способность написать эссе, которое конкурсное жюри не смогло отличить от написанного человеком

Появление у модели качественно новых способностей



- GPT-3: 11/Jun/2020, контекст 1536 слов (3 страницы)
- 175 млрд. параметров, корпус 500 млрд. токенов
- способность делать перевод на другие языки,
- решать логические и математические задачи,
- генерировать программный код по описанию

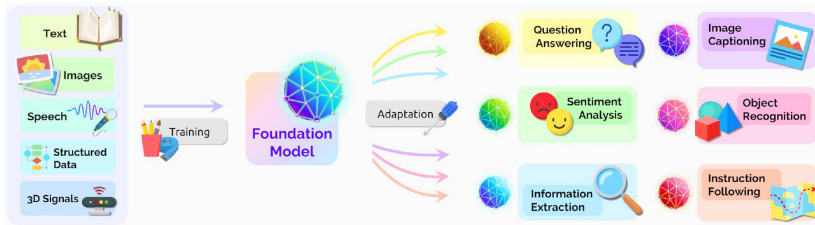
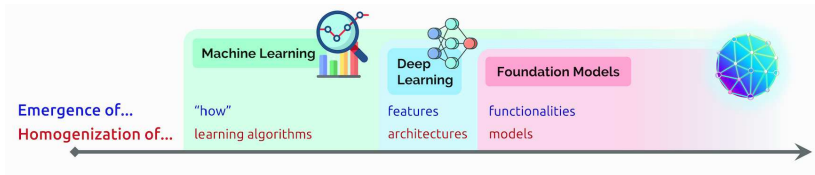
Появление у модели качественно новых способностей



- GPT-4: 14/Mar/2023, контекст 24 000 слов (48 страниц)
- >1 трл. параметров, корпус >1Tb
- способность описывать и анализировать изображения,
- реагировать на подсказки вроде «Let's think step by step»,
- решать качественные физические задачи по картинке

Гомогенизация векторных моделей (Foundation Models)

Обучаемая векторизация данных — глобальный тренд AI/ML



R. Bommasani et al. (Center for Research on Foundation Models, Stanford University)
 On the opportunities and risks of foundation models // CoRR, 20 August 2021.

Начинаем с постановки задачи: ДНК (дано, найти, критерий)









П.Домингос (2015) ошибался, что «верховный алгоритм» (AGI) появится как объединение подходов основных научных школ

Ю.И.Журавлёв в *алгоритмах вычисления оценок* (1971) уже сделал это, но это не привело к появлению AGI

Теперь ясно, что слагаемые успеха — совсем другие:

- 1 обучаемая векторизация данных → глубокое обучение
- 2 большие данные + большие модели + нейропроцессоры
- 3 вычислительные методы оптимизации без переобучения
- 4 обучение большой языковой модели → эмерджентность

Полезные ссылки и литература

-  *Николенко С.* Машинное обучение: основы, 2025.
-  *Николенко С., Кадурын А., Архангельская Е.* Глубокое обучение: основы, 2024.
-  *Кристофер Бишоп, Хью Бишоп.* Глубокое обучение: принципы и концепции, 2025.
-  *Мэрфи К.П.* Вероятностное машинное обучение. Введение, 2022.
-  *Мэрфи К.П.* Вероятностное машинное обучение. Дополнительные темы: основания, вывод, 2024.
-  *Мэрфи К.П.* Вероятностное машинное обучение. Дополнительные темы: предсказание, порождение, обнаружение, действие, 2024.
-  *Дайзенрот М.П., Альдо Фейзал А., Чен Сунь Он.* Математика в машинном обучении, 2024.
-  *Уилке К.* Основы визуализации данных: пособие по эффективной и убедительной подаче информации, 2024.
-  *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.
-  *Визильтер Ю. В.* От слабого ИИ к общему универсальному интеллекту (обзор тенденций 2020-2023). Семинар РАИИ и ФИЦ ИУ