Конгресс «Молекулярная диагностика и биобезопасность-2024», Москва, 16-17 апреля 2024
Секция «Эпидемиологический анализ и прогнозирование в условиях цифровой трансформации»

Эволюция идей искусственного интеллекта: история развития и связь с практикой

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,
зав. кафедрой математических методов прогнозирования МГУ,
зав. кафедрой интеллектуальных систем МФТИ,
voron@mlsa-iai.ru

Методология эмпирической индукции

От дедуктивного метода познания к индуктивному:
«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)

Таблицы открытия (*выборки эмпирических данных*):
множества объектов x, у которых целевое свойство $y(x)$

- присутствует или отсутствует (*классификация*)
- принимает различные числовые значения (*регрессия*)

Фрэнсис Бэкон. Новый органон. 1620.

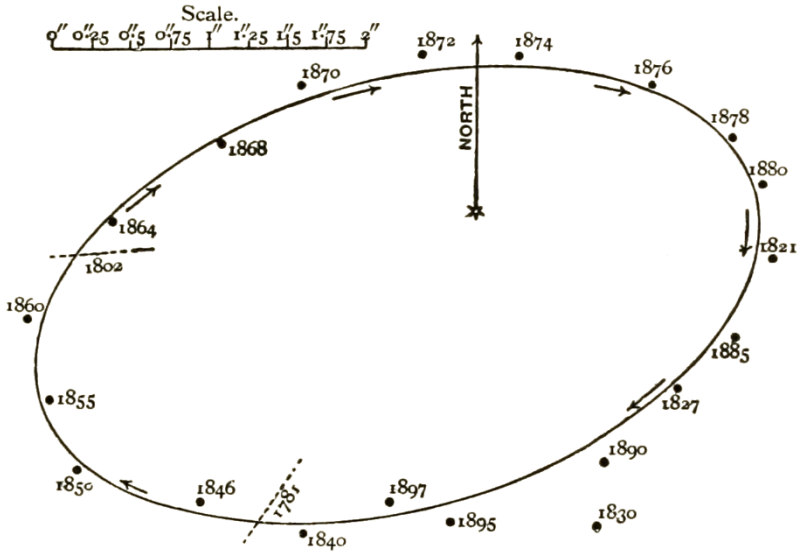
Задача проведения функции через точки

Предсказание свойства $y(x)$ по признакам $f_j(x)$,
(линейной) модели $a(x, w)$ с параметрами w:

$$a(x, w) = \sum_j w_j f_j(x)$$

Метод наименьших квадратов (Гаусс, 1795):

$$\sum_{(x,y)} (a(x, w) - y)^2 \to \min_w$$



Карл Фридрих Гаусс
(1777--1855)

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

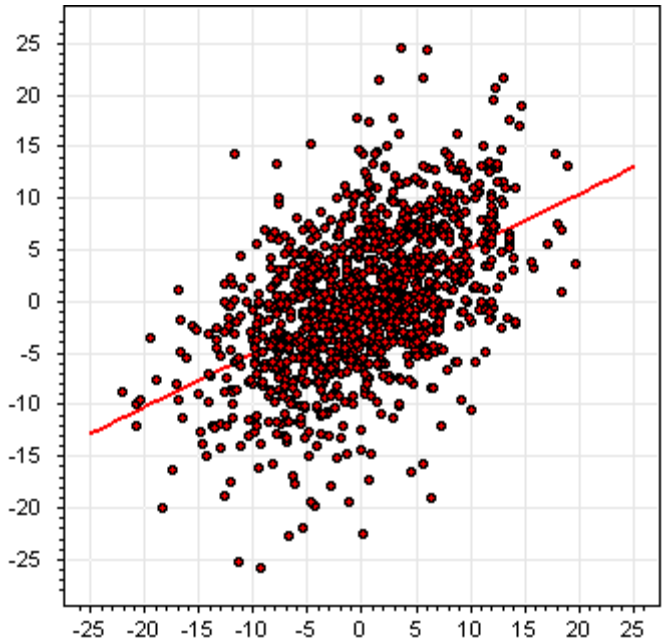
C.F.Gauss. Theory of the motion of the heavenly bodies moving about the Sun in conic sections. 1809.

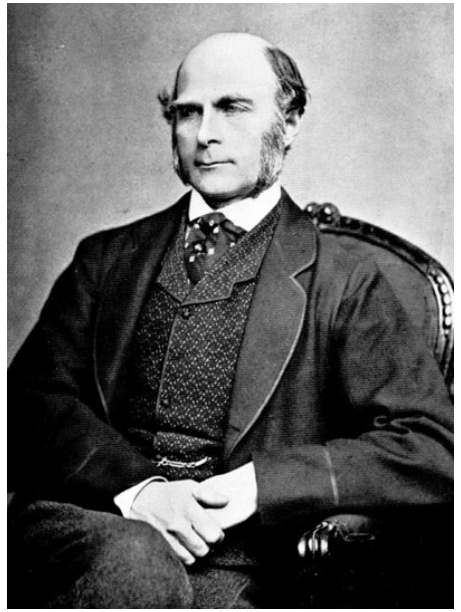
Задача восстановления регрессии

Исследование наследственности роста (Гальтон, 1886).

Δ — отклонение роста от среднего в популяции

Зависимость (линейная?) Δ взрослого сына от Δ отца:





Фрэнсис Гальтон
(1822--1911)

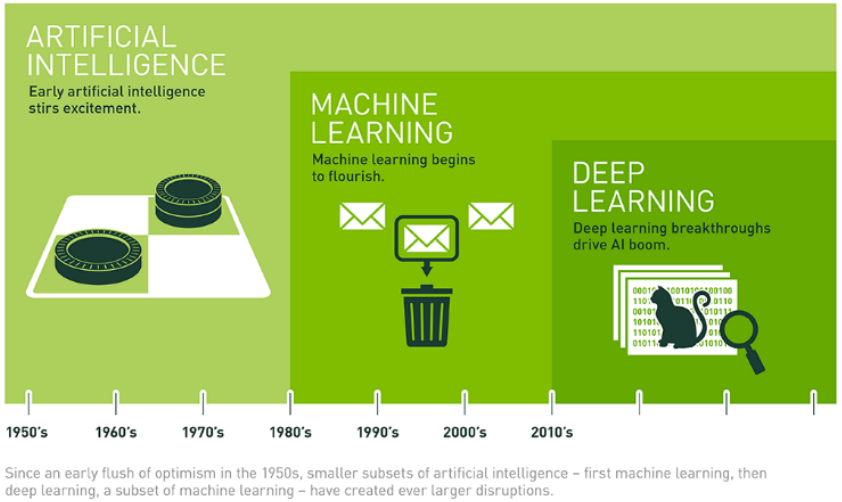
Двойной смысл термина «регрессия»:

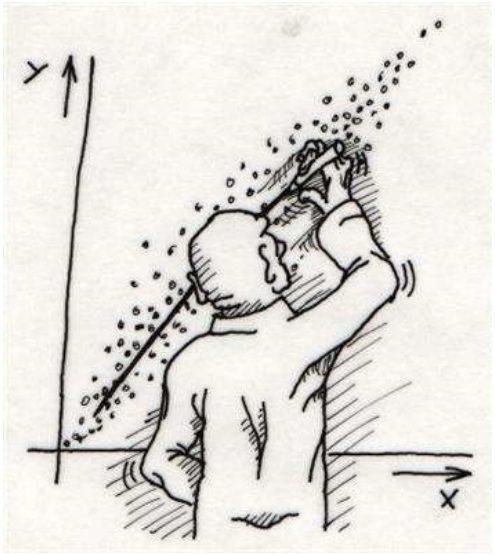
- регрессия (роста)
  к среднему в популяции
- *необычный «обратный» ход исследования: сначала данные, потом модель*

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

Машинное обучение (Machine Learning, ML)

- одна из ключевых информационных технологий будущего
- наиболее успешное направление ИИ, вытеснившее экспертные системы и инженерию знаний





- проведение функции через заданные точки в сложно устроенных пространствах
- математическое моделирование в условиях, когда знаний мало, данных много
- тысячи различных методов и алгоритмов
- более 100 000 научных публикаций в год

Задачи машинного обучения с учителем

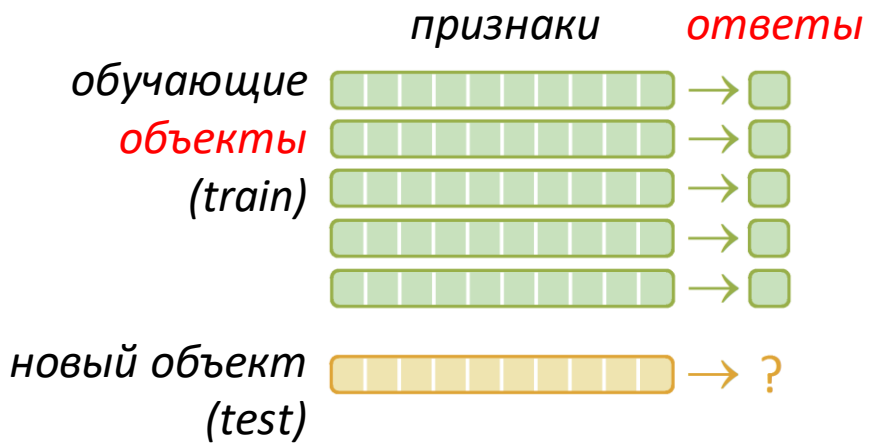
Этап №1 – обучение с учителем

- **На входе:**
  данные – выборка прецедентов «*объект* → *ответ*»,
  каждый объект описывается набором *признаков*
- **На выходе:**
  модель, предсказывающая ответ по объекту

Если нет данных,
то нет
и машинного
обучения

Этап №2 – применение

- **На входе:**
  данные – новый объект
- **На выходе:**
  предсказание ответа на новом объекте



Примеры задач машинного обучения

- **Медицинская диагностика:**
  объект – данные о пациенте на текущий момент
  ответ – диагноз / решения о мероприятиях

- **Предсказание инфицирования в результате контакта:**
  объект – данные с *носимого устройства* (http://amuleit.ru)
  ответ – вероятность передачи инфекции

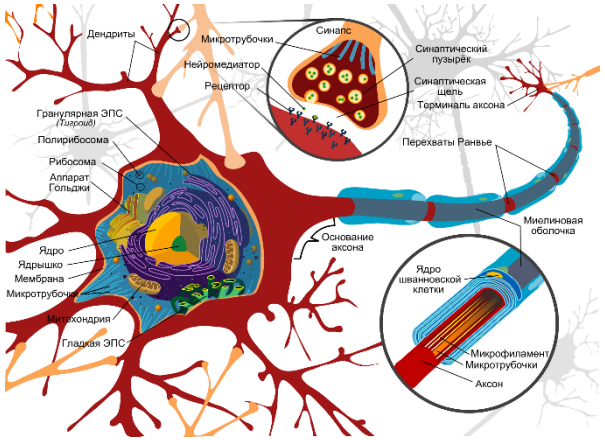
- **Предсказание инфицирования по множеству контактов:**
  объект – данные о контактах индивида в интервале времени
  ответ – вероятность инфицирования

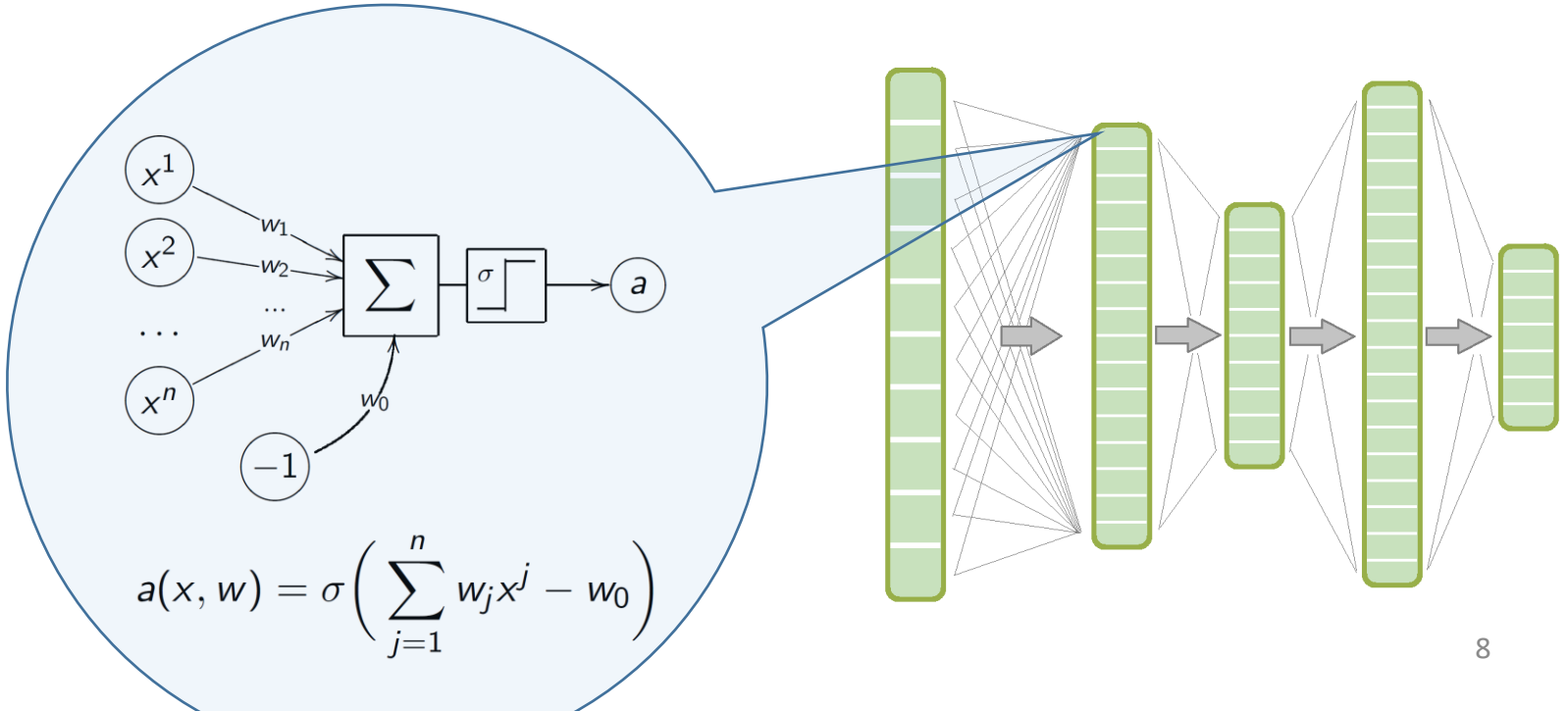
Модели искусственных нейронных сетей

На каждом слое сети вектор объекта преобразуется в новый вектор

Каждое преобразование (нейрон) – линейная модель $a(x, w)$

Веса w являются обучаемыми параметрами модели

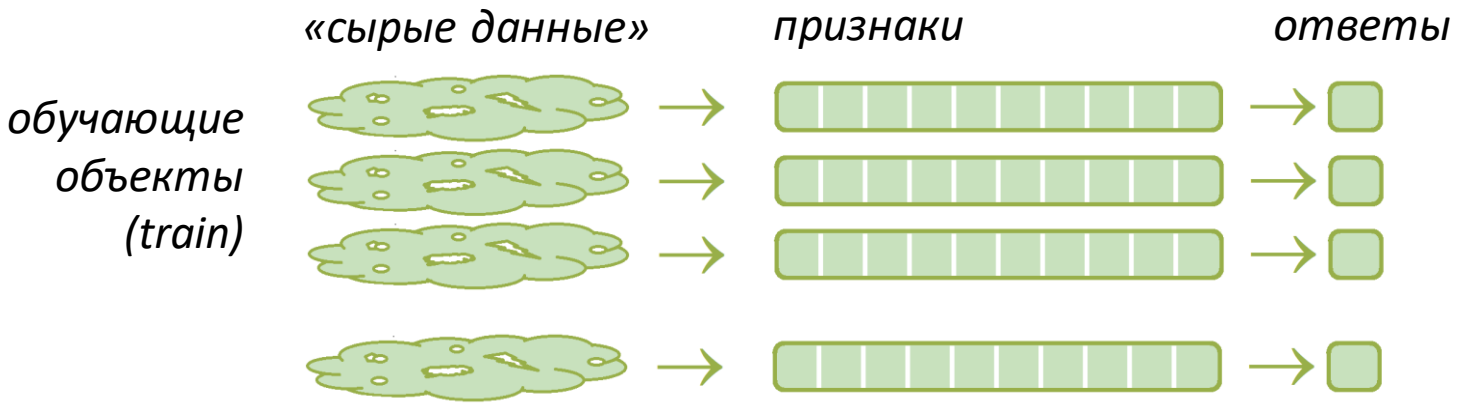




Модели глубоких нейронных сетей

Вход: сложно структурированные «сырые» данные объектов

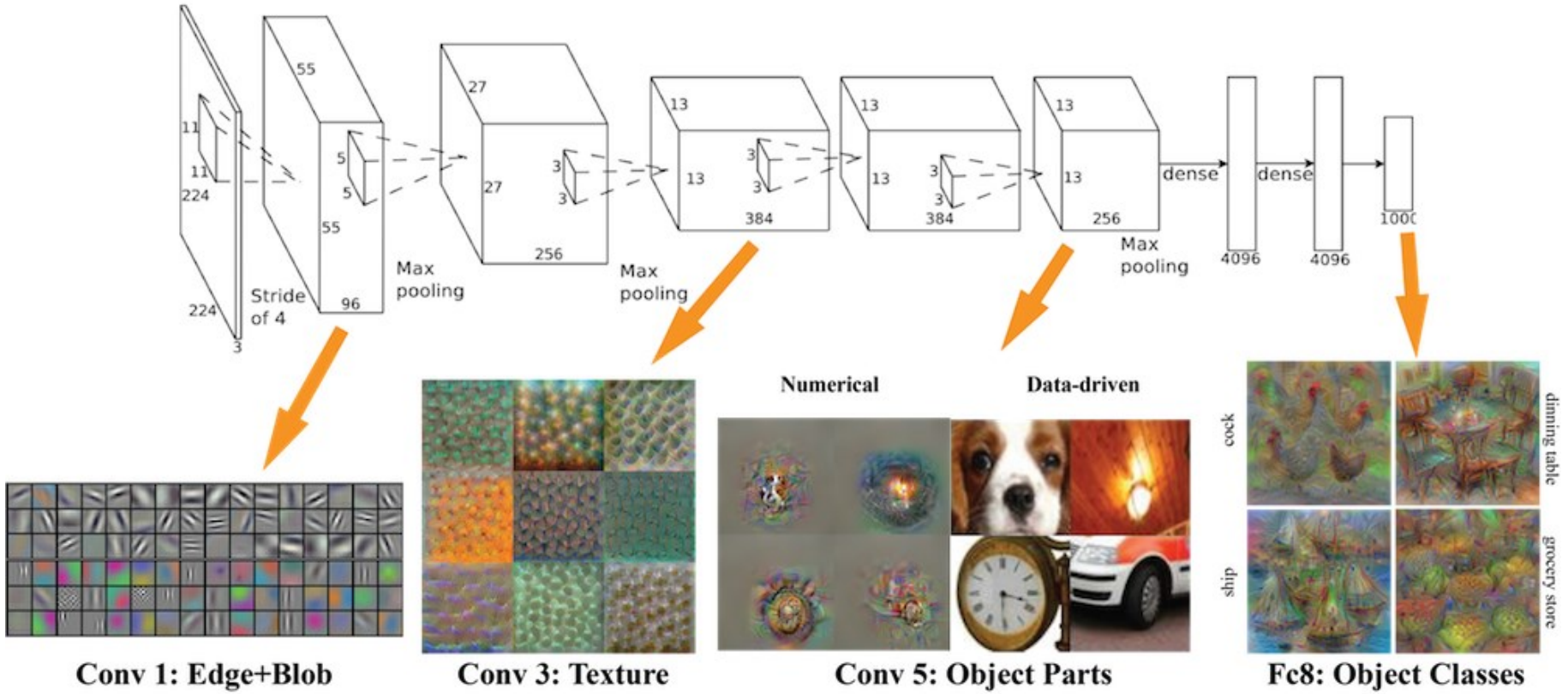
Выход: векторные признаковые представления объектов, затем ответы



Deep Learning – это всего лишь обучаемая векторизация сложных объектов

Примеры сложно структурированных объектов:
изображения, видео, временные ряды, тексты, транзакции, графы, ...

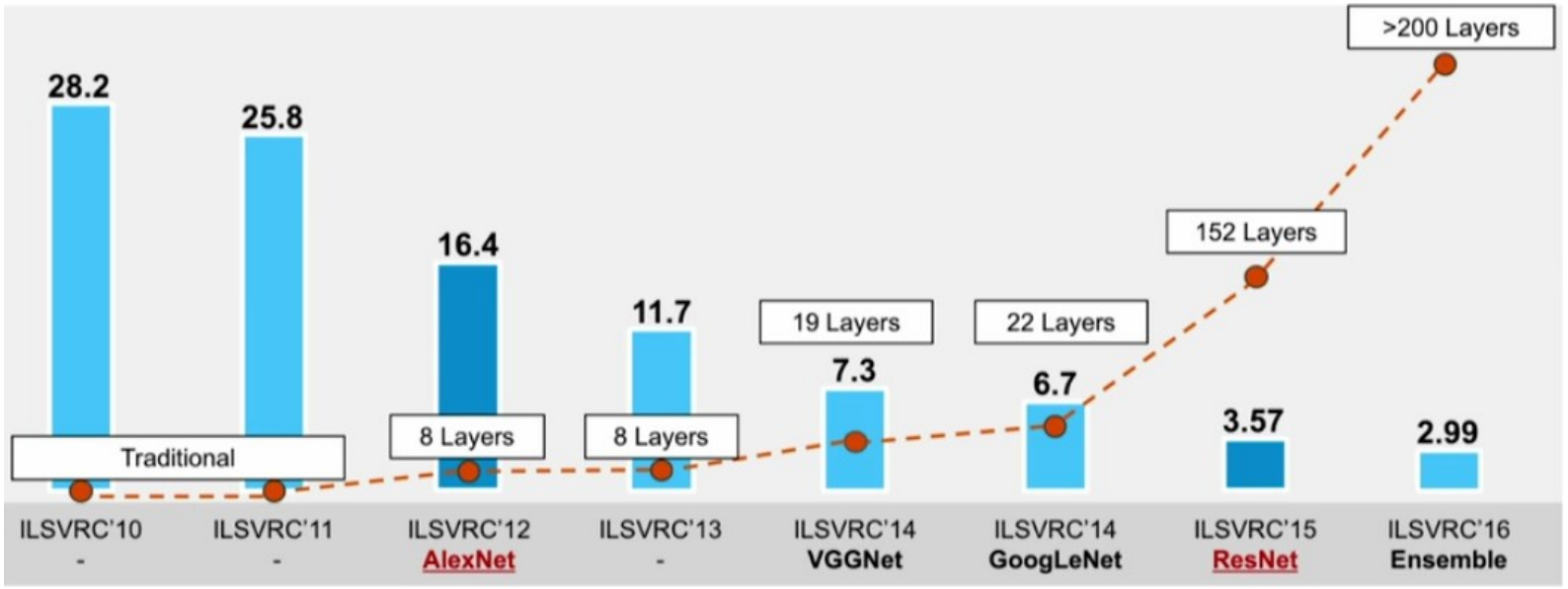
Глубокие свёрточные нейронные сети
для классификации объектов на изображениях



Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. 2012.

Роль больших данных

ImageNet: открытая выборка 14M изображений, 20K категорий



Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

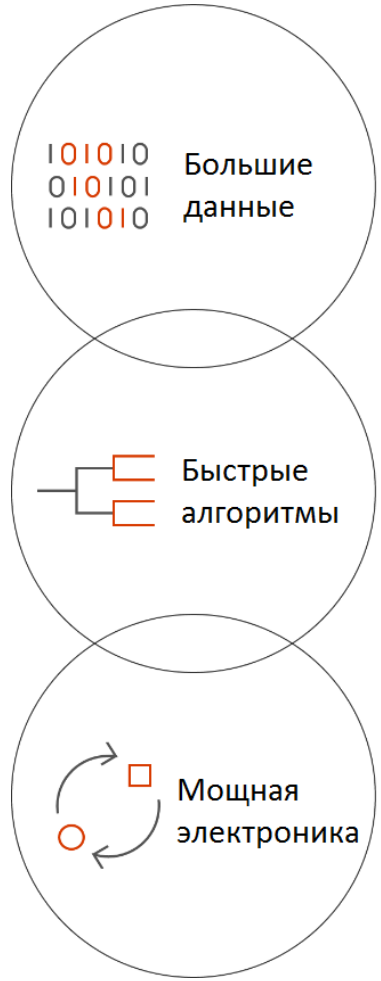
Li Fei-Fei et al. ImageNet: A large-scale hierarchical image database. 2009.
Li Fei-Fei et al. Construction and analysis of a large scale image ontology. 2009.

Три составляющих успеха Deep Learning

- Повсеместное применение компьютерных технологий
  → *накопление больших выборок данных*
  в частности, ImageNet

- Развитие математических методов и алгоритмов
  → *накопление критической массы опыта*
  методы оптимизации для больших размерностей

- Достижения микроэлектроники
  → *рост вычислительных мощностей, закон Мура*
  в частности, графические ускорители (GPU)



Машинное обучение – это оптимизация

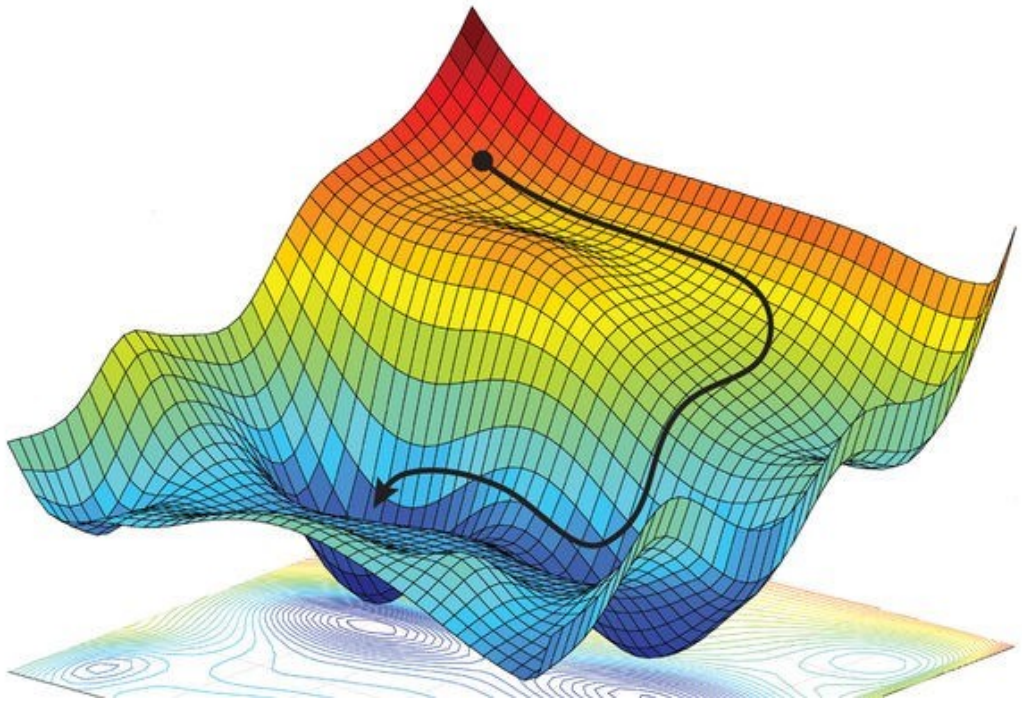
x – вектор объекта обучающей выборки

$a(x, w)$ – предсказательная модель

w – параметры модели

$\mathrm{Loss}(x, w)$ – функция потерь

$Q(w)$ – критерий качества модели



Задача обучения параметров модели:

$$Q(w) = \sum_{x} \mathrm{Loss}(x, w) \quad \to \min$$

Способ решения – численные методы оптимизации

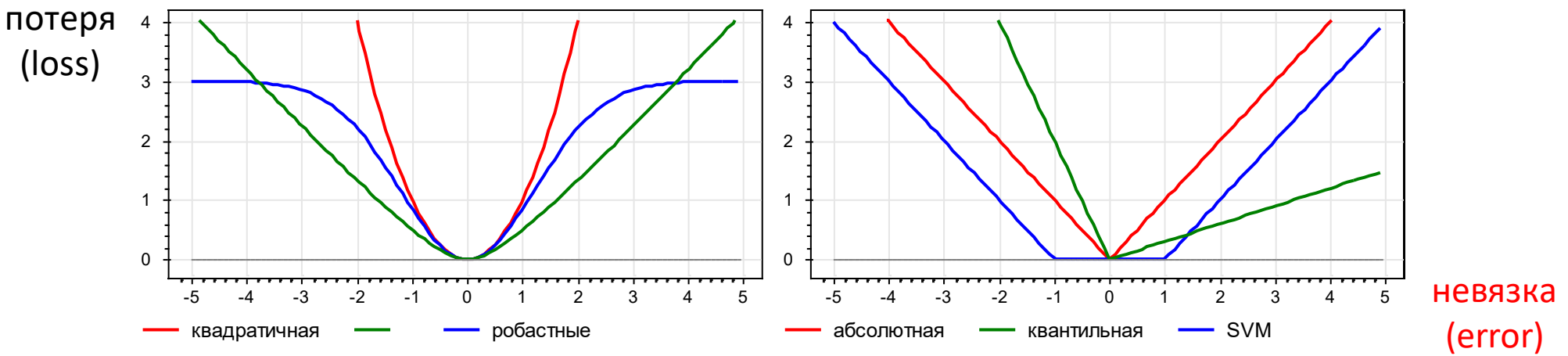
Обучение с учителем (supervised learning): восстановление регрессии (regression)

x — вектор объекта обучающей выборки, y — числовой ответ

$a(x, w)$ — модель регрессии с параметрами w

Например, $a(x, w) = \sum_j w_j x_j$ — линейная модель регрессии

$\text{Loss}(x, w) = (a(x, w) - y)^2$ — квадратичная функция потерь



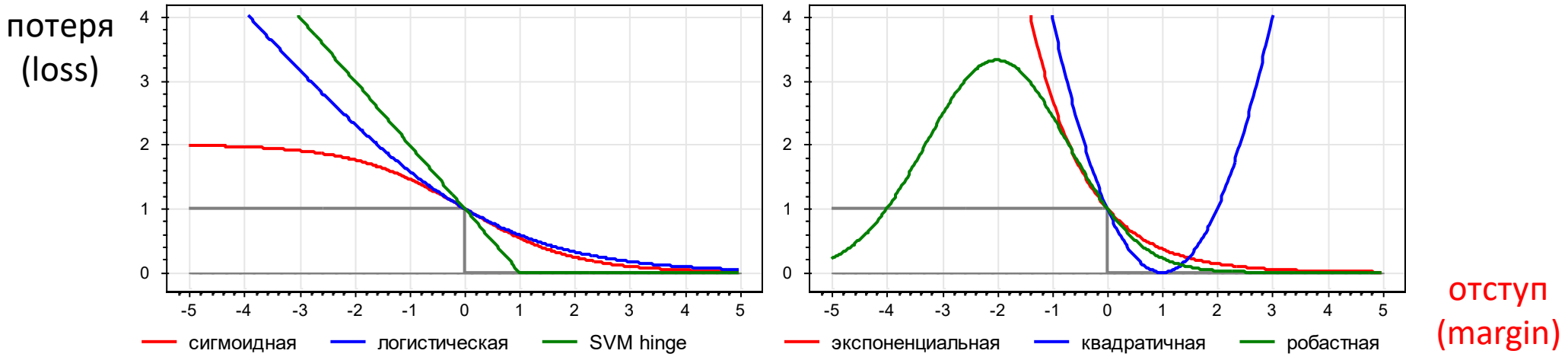
Обучение с учителем (supervised learning): классификация (classification)

x — вектор объекта обучающей выборки, y — ответ ($+1$ или -1)

$a(x, w)$ — модель классификации с параметрами w

Например, $a(x, w) = \text{sign}\left(\sum_j w_j x_j\right)$ — линейная модель

$\text{Loss}(x, w) = \max\left(0, 1 - y\sum_j w_j x_j\right)$ — функция потерь SVM hinge



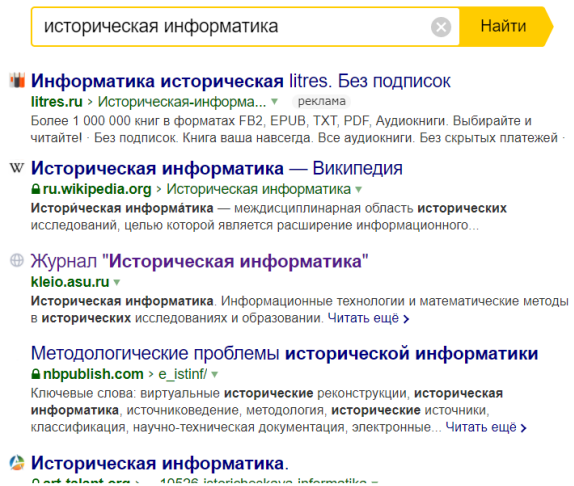
Обучение с учителем (supervised learning): обучение ранжированию (learning to rank)

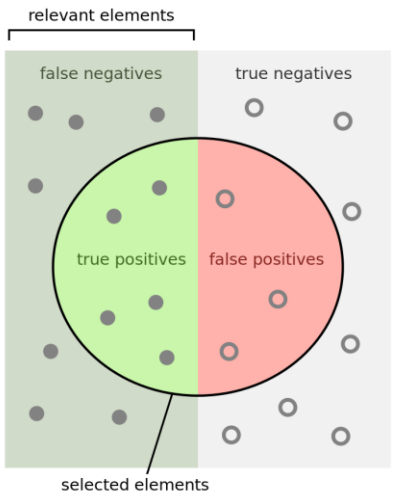
x — вектор пары «запрос-документ», y — оценка релевантности

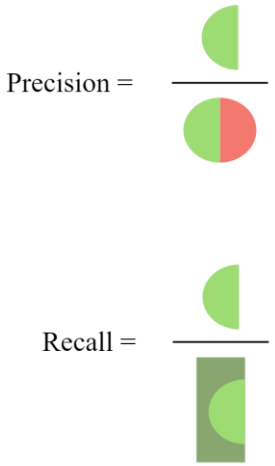
$a(x, w)$ — модель ранжирования документов по запросу, параметр w

Например, $a(x, w) = \sum_j w_j x_j$ — линейная модель

$$\text{Loss}(x, x', w) = \max\Big(0, 1 - [y > y'](a(x, w) - a(x', w))\Big)$$







не только поиск,
но и любые задачи, где
человеку удобно
принимать решения,
выбирая один из вариантов

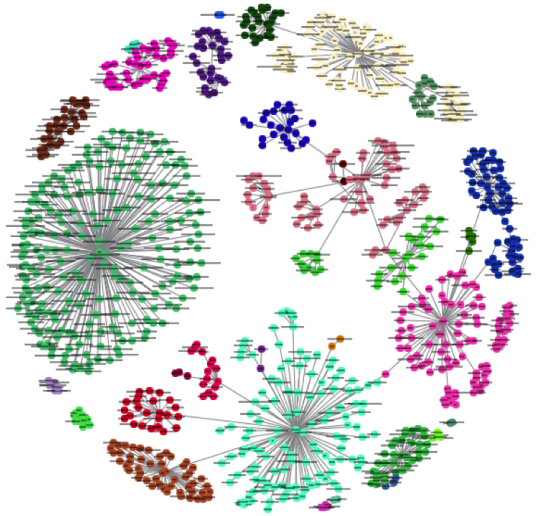
Обучение без учителя (unsupervised learning): кластеризация (clustering)

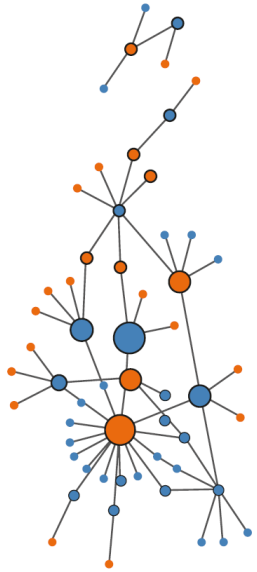
x — вектор объекта обучающей выборки, ответы не задаются

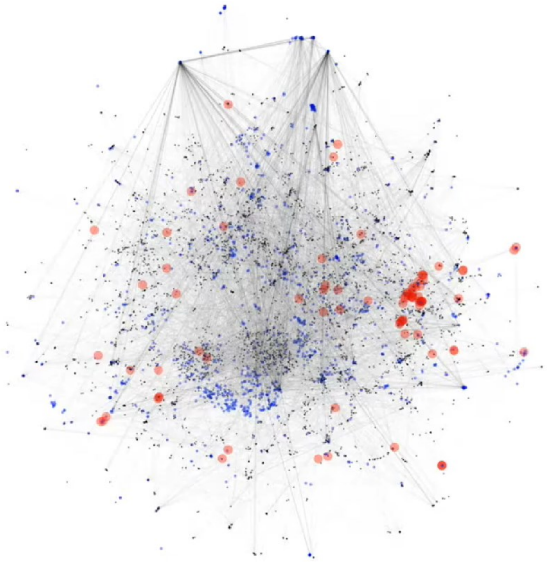
$a(x, w)$ — кластер, ближайший к x

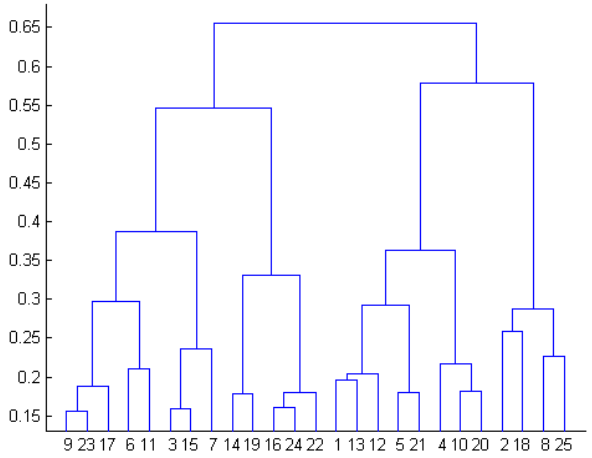
$w = \{c_1, \dots, c_K\}$ — векторы центров всех кластеров

$\text{Loss}(x, w) = \min_k \|x - c_k\|$ — расстояние до ближайшего кластера









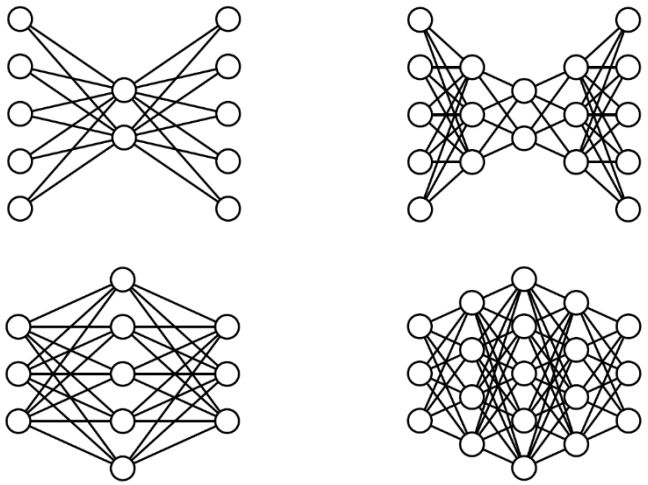
Обучение без учителя (unsupervised learning): векторизация, автокодировка (autoencoder)

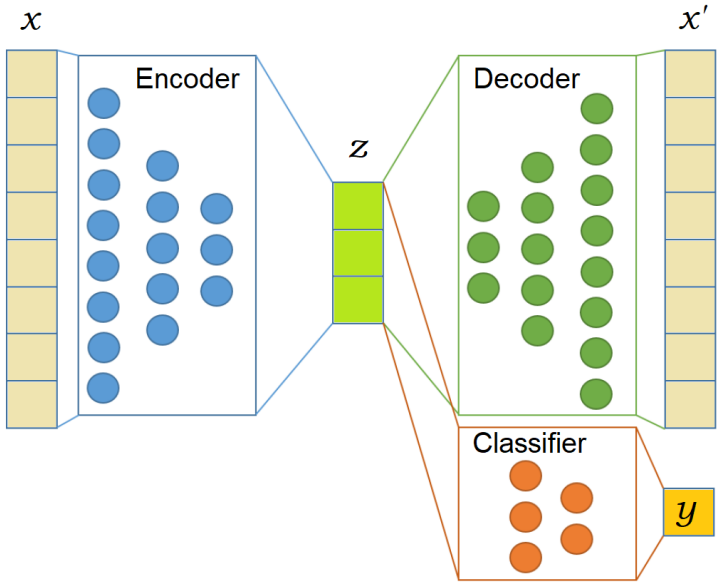
x — описание объекта обучающей выборки, ответов не дано

$z = f(x, w)$ — модель кодирования x в векторное представление z

$x' = g(z, w')$ — модель декодирования z в реконструкцию x'

$\text{Loss}(x, w) = \|g(f(x, w), w') - x\|$ — точность реконструкции объекта





обучаемая векторизация сложных объектов

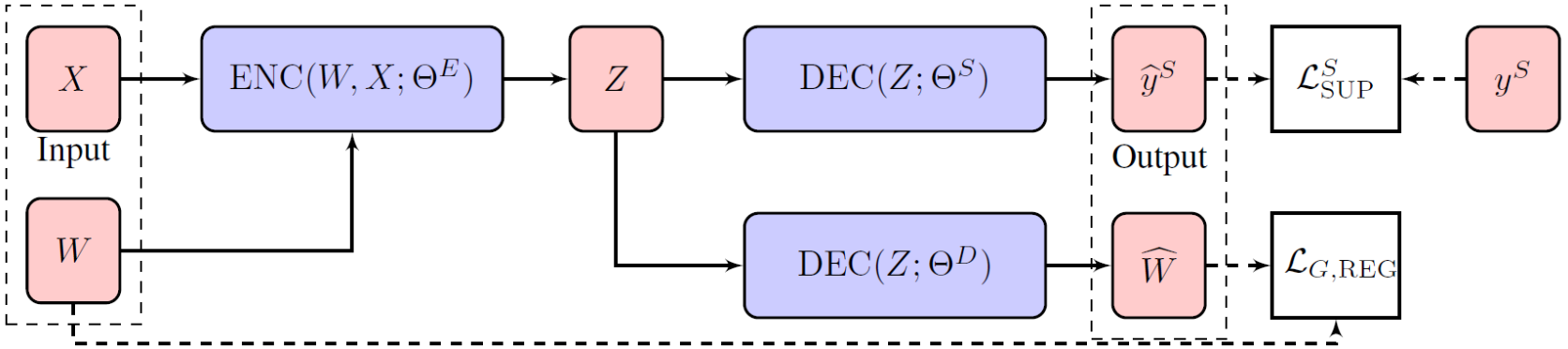
Частичное обучение (semi-supervised learning): векторизация графов (graph embeddings)

$x; (x, x')$ — данные об объектах и взаимодействиях между объектами

$z = f(x, \theta^E)$ — модель векторизации объектов x (вершин графа)

$x' = g(z, \theta^D)$ — модель декодирования z в реконструкцию x'

$\text{Loss}(x, w) = \|g(f(x, \theta^E), \theta^D) - x\| + \tau L^S_{\text{SUP}}(x, \theta^S)$ — два критерия



обучаемая векторизация сложных объектов по данным об их взаимодействиях

T.Mikolov et al. Efficient estimation of word representations in vector space, 2013.

I.Chami et al. Machine learning on graphs: a model and comprehensive taxonomy. 2020.

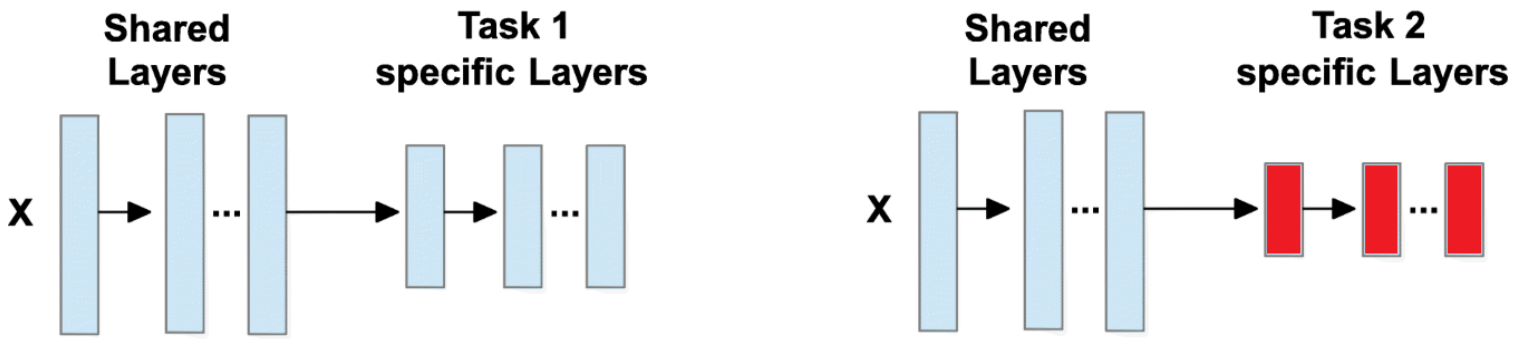
Перенос обучения (transfer learning), предобучение модели векторизации

$z = f(x, w)$ — модель векторизации, универсальная для многих задач

$y = g(z, w')$ — часть модели, специфичная для своей задачи

$\min\limits_{w,\, w'}: \sum_x \text{Loss}_1(g_1(f(x, w), w'))$ — обучение по большим данным

$\min\limits_{w'}: \sum_{x'} \text{Loss}_2(g_2(f(x', w), w'))$ — обучение по своим данным

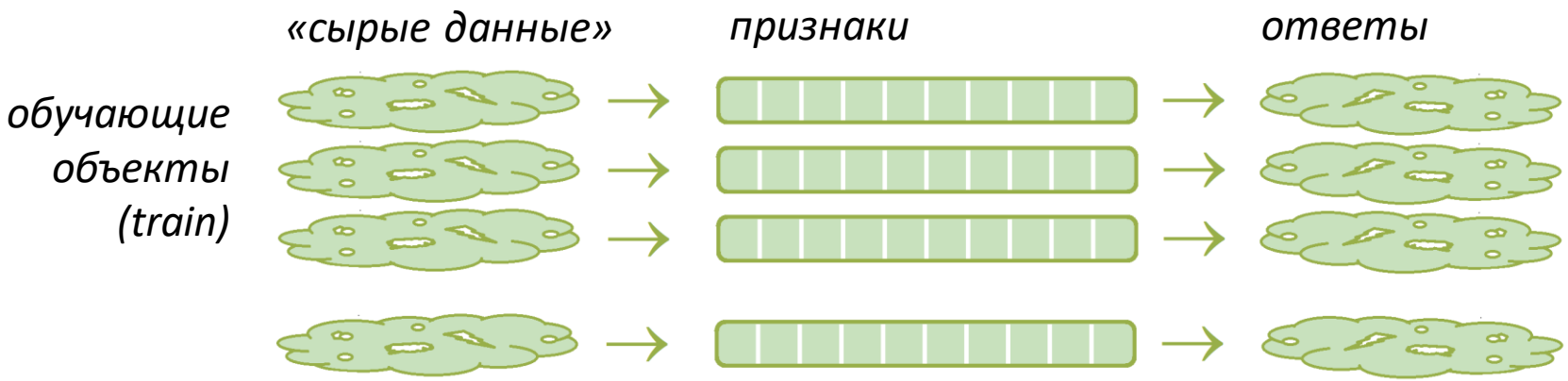


Sinno Jialin Pan, Qiang Yang. A Survey on Transfer Learning. 2009

Нейронные сети для синтеза объектов

Вход: сложно структурированные объекты

Выход: сложно структурированные ответы



Примеры: синтез изображений, перенос стиля, распознавание речи, машинный перевод, суммаризация текстов, диалог с пользователем

Модели: seq2seq, CNN, RNN, LSTM, GAN, BERT, GPT и др.

Генеративная состязательная сеть (GAN)

$x = g(z, w)$ — модель генерации реалистичного объекты x из шума z

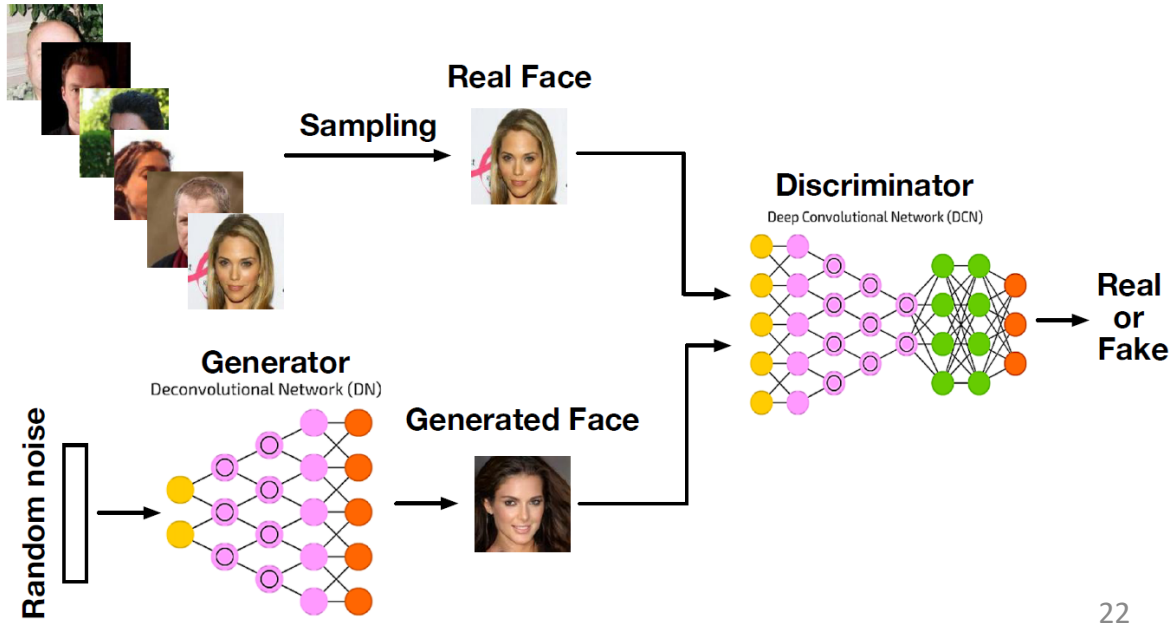
$f(x, w')$ — модель классификации x «реальный/сгенерированный»

$\min_{w} \max_{w'} \sum_x \ln f(x, w') + \ln\left(1 - f(g(z, w), w')\right)$ — совместное обучение

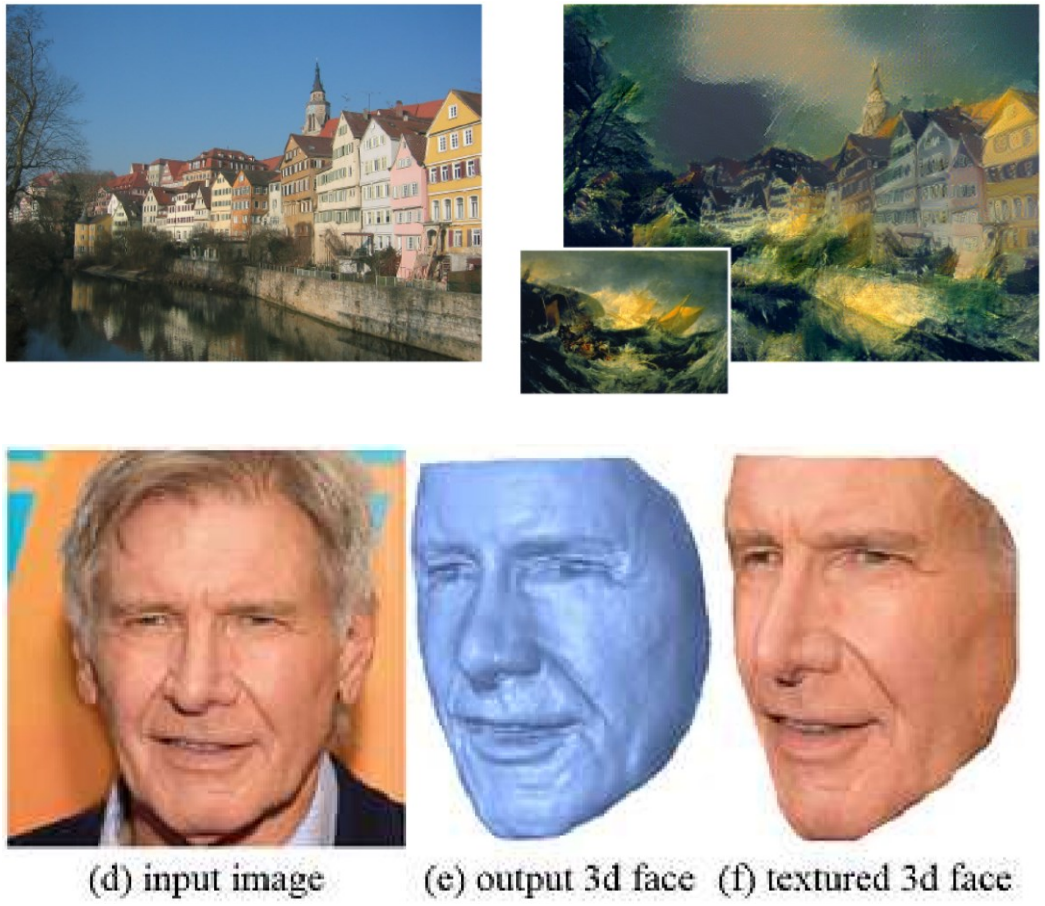
Antonia Creswell et al. Generative Adversarial Networks: an overview. 2017.

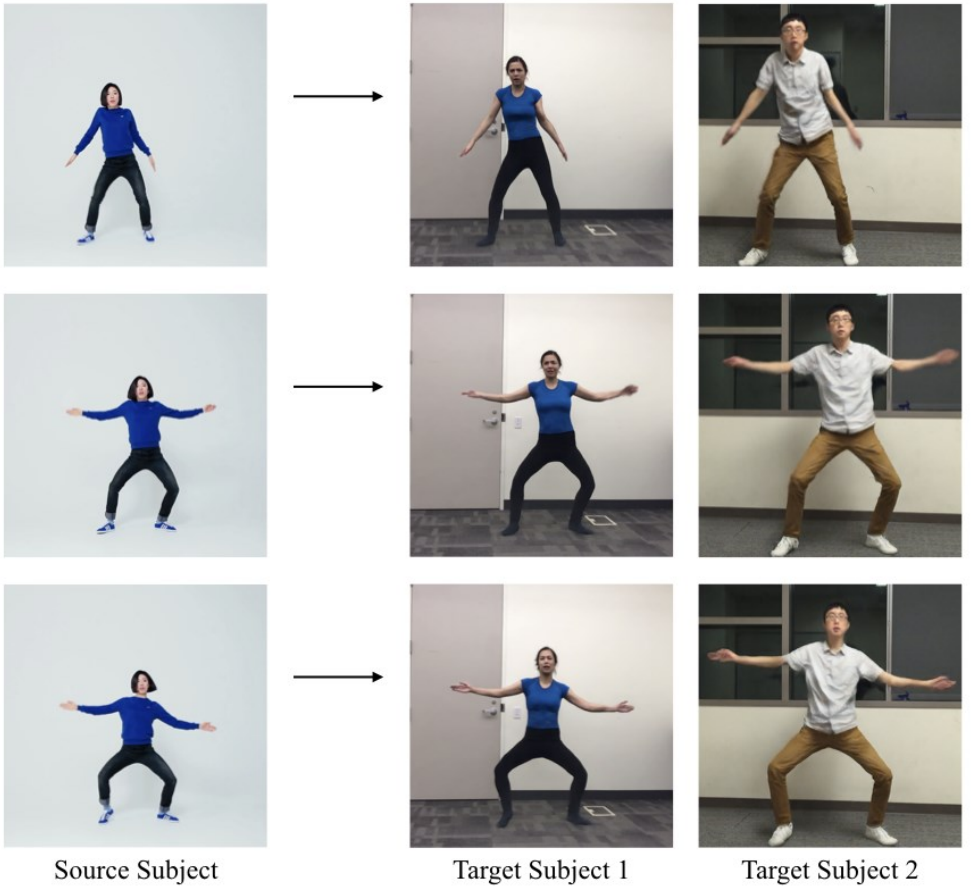
Zhengwei Wang et al. Generative Adversarial Networks: a survey and taxonomy. 2019.

Chris Nicholson. A Beginner's Guide to Generative Adversarial Networks. 2019.



Синтез изображений и видео





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

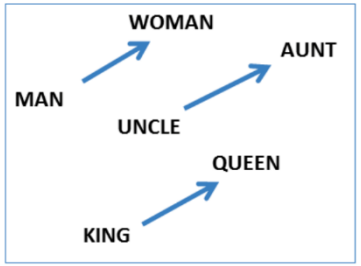
Эволюция подходов в обработке текстов

Декомпозиция задач по уровням «пирамиды NLP»

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

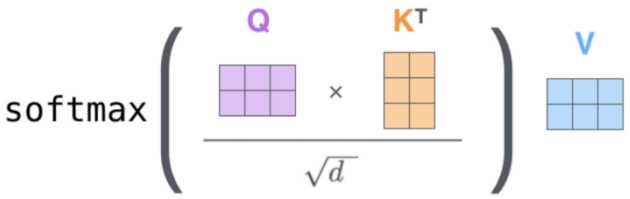
Модели векторизации слов (эмбедингов)

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

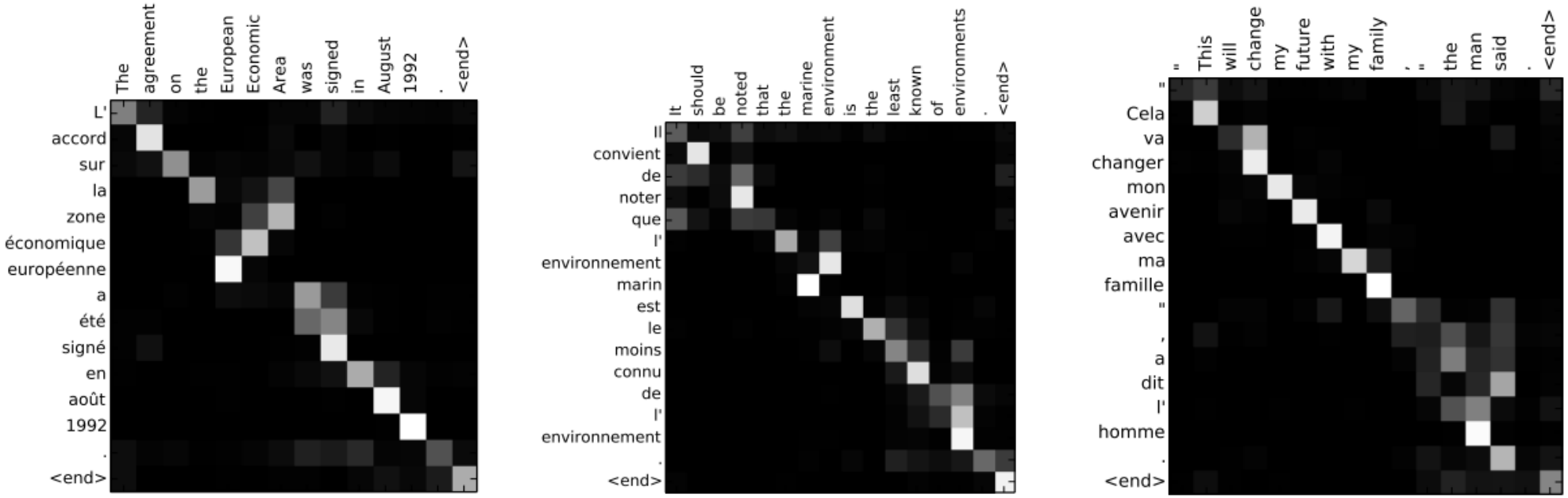


Нейросетевые модели контекстной векторизации

- рекуррентные нейронные сети: LSTM, GRU, …
- «end-to-end» модели внимания и трансформеры:
  машинный перевод [2017], BERT [2018], GPT-4 [2023], …



Модели внимания: машинный перевод

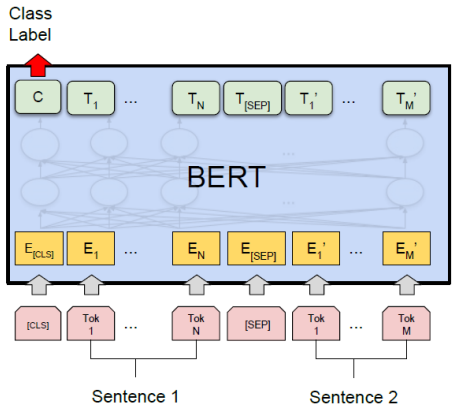


Интерпретация моделей внимания: *матрица семантического сходства* A[t,i] показывает, на какие слова x[i] входного текста модель обращает внимание, когда генерирует слово перевода y[t]

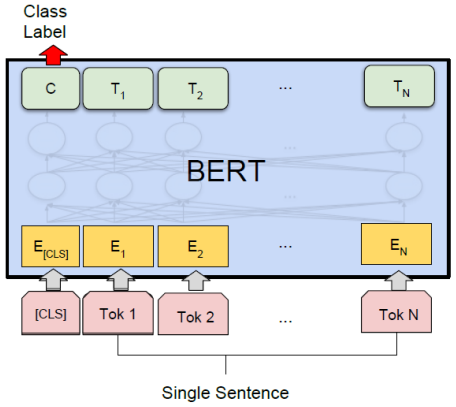
Bahdanau et al. Neural machine translation by jointly learning to align and translate. 2015

Трансформеры: нейросетевые модели языка

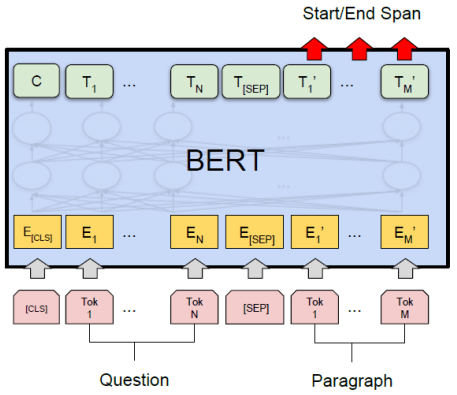
- Обучаются векторизовать и предсказывать слова по контексту
- Обучаются по терабайтам текстов, «они видели в языке всё»
- Мультиязычны: обучаются на десятках языков
- Мультизадачны: для каждой новой задачи NLP/NLU достаточно предобученной модели или дообучения на небольшой выборке



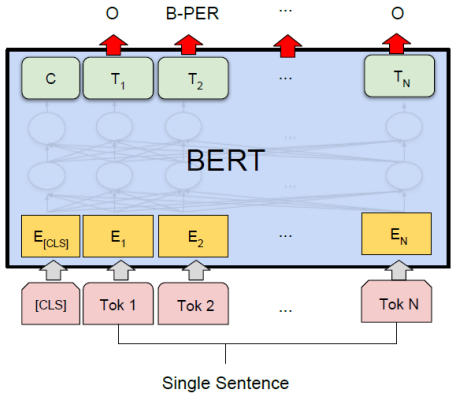
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG



(b) Single Sentence Classification Tasks: SST-2, CoLA



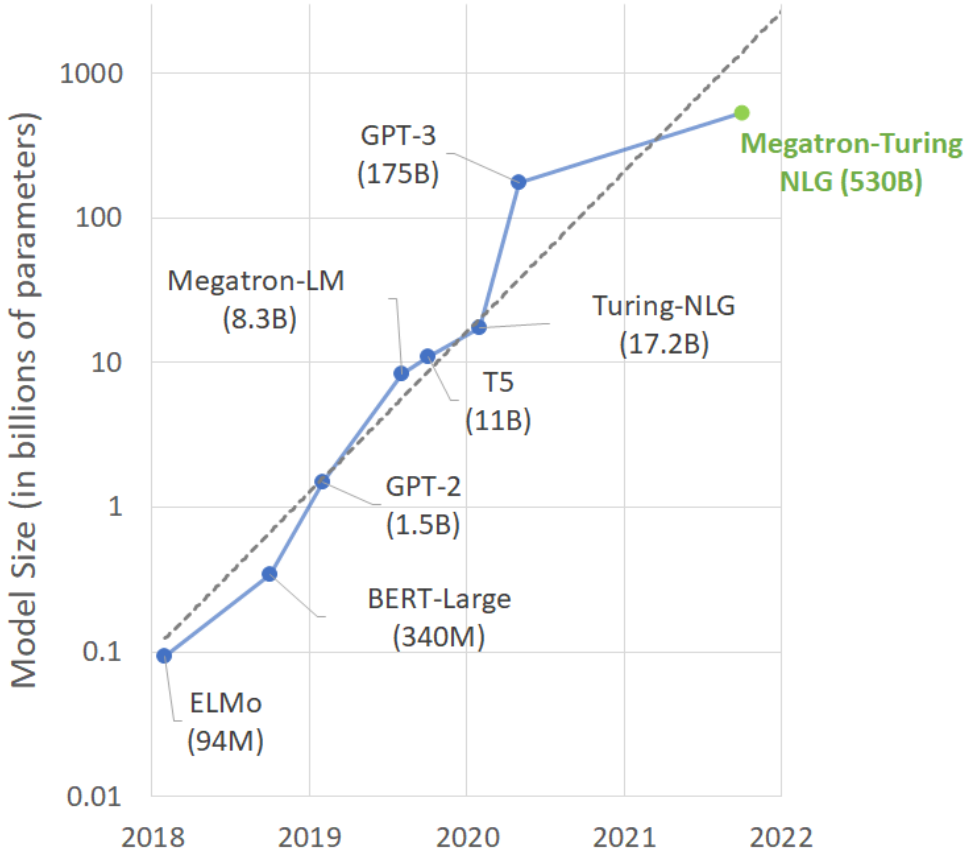
(c) Question Answering Tasks: SQuAD v1.1

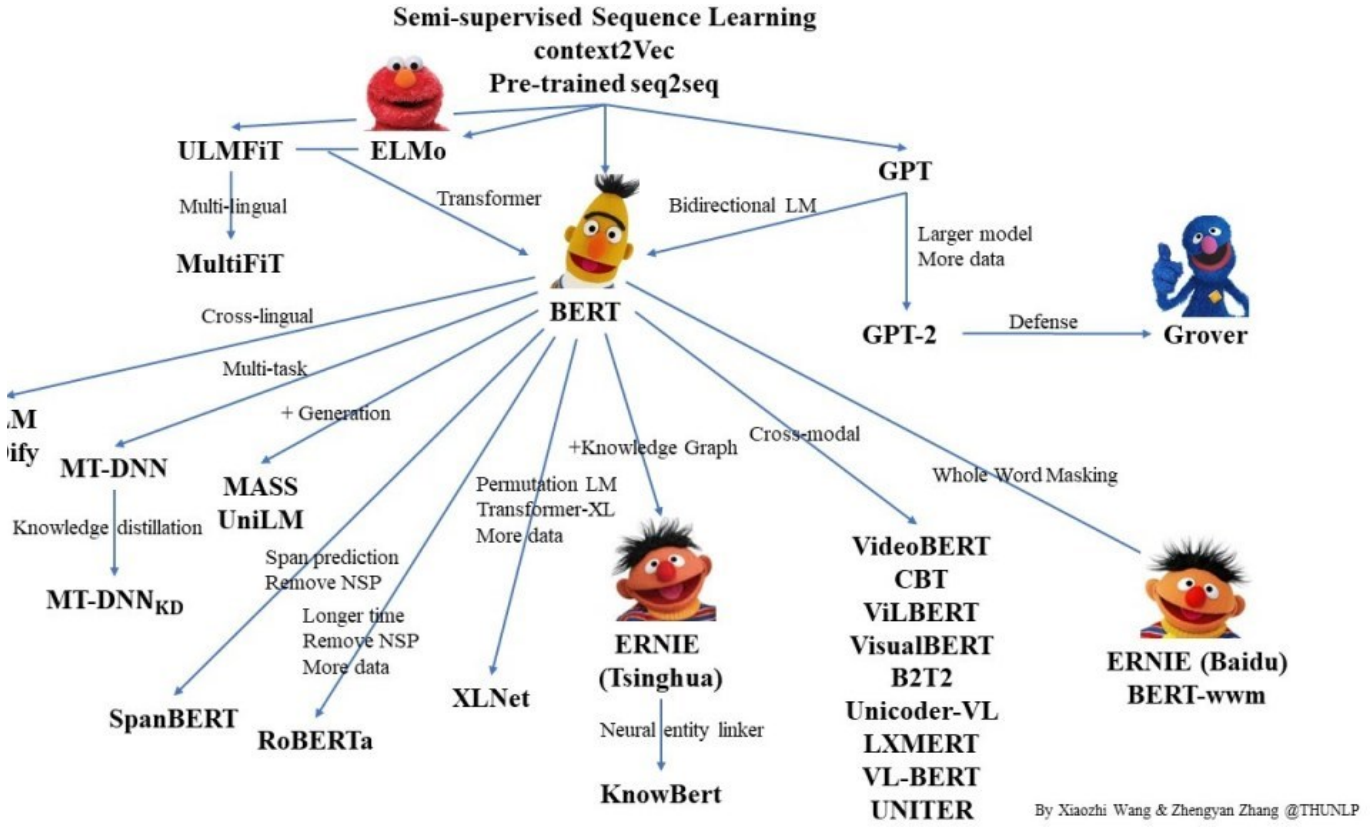


(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Трансформеры: нейросетевые модели языка

Рост числа параметров нейросетевых трансформерных моделей языка





Проблески общего искусственного интеллекта

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

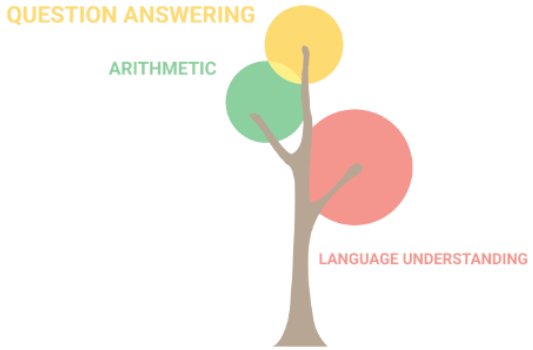
Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research *(27 March 2023)*

Новые способности модели, не закладывавшиеся при обучении:

- объяснять свои ответы, перефразировать, переводить на другие языки
- реферировать, генерировать планы, сценарии, шаблоны
- 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

Новые (эмерджентные) способности модели

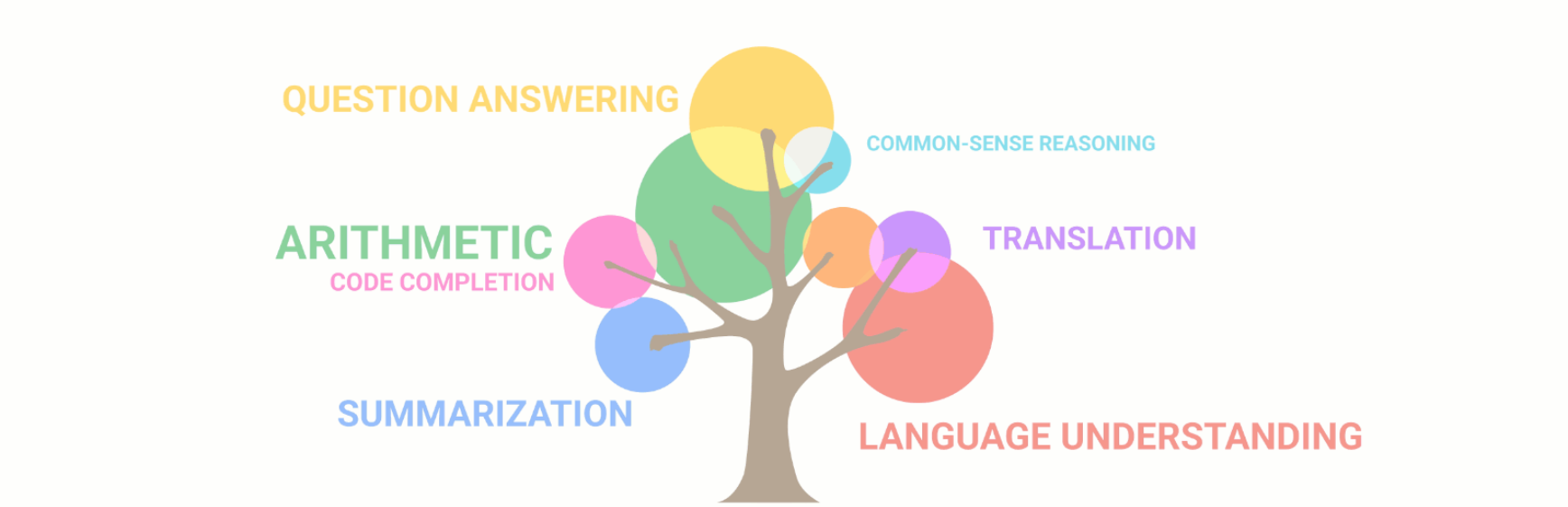


GPT-2: 14-Feb-2019

1,5 млрд. параметров, корпус 10 млрд. токенов (40Gb), контекст 768 слов (1,5 стр.)

- способность написать эссе, которое конкурсное жюри не смогло отличить от написанного человеком

Новые (эмерджентные) способности модели

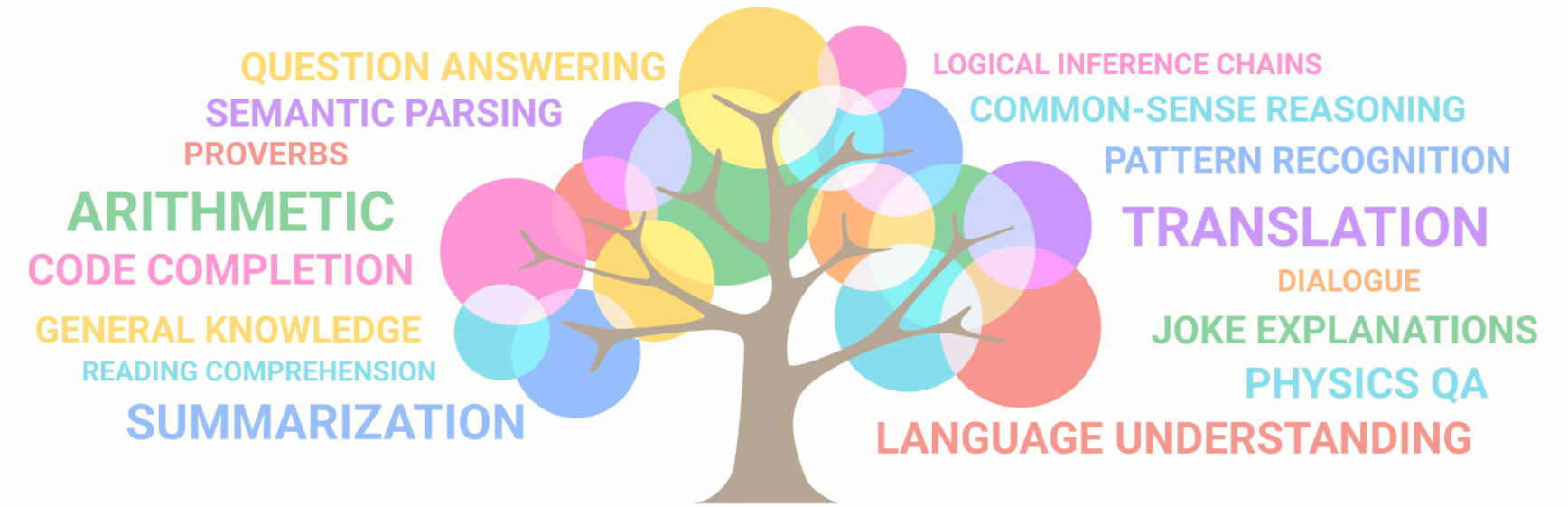


GPT-3: 11-Jun-2020

175 млрд. параметров, корпус 500 млрд. токенов, контекст 1536 слов (3 стр.)

- способность делать перевод на другие языки
- способность решать логические и простейшие математические задачи
- способность генерировать программный код по текстовому описанию

Новые (эмерджентные) способности модели



GPT-4: 14-Mar-2023

>1 трл. параметров, корпус >1Tb, контекст 24 000 слов (48 страниц)

- способность описывать и анализировать изображения
- способность реагировать на подсказки вроде «Let's think step by step»
- способность решать качественные физические задачи по картинке

Возможности и угрозы

Чаты GPT уже способны помогать с рутинно-творческой работой:

- генерировать документы или сайты по техническому заданию
- в том числе медицинские, юридические документы по шаблонам
- искать и структурировать профессиональную информацию
- делать обзоры, рефераты, сводки на разных языках
- генерировать программный код по описанию
- обсуждать новости, поддерживать разговор по теме
- разговаривать с детьми с учётом возрастных особенностей
- выполнять функции воспитателя, учителя, наставника
- оказывать психологическую помощь

Возможности и угрозы

Чаты GPT способны (непреднамеренно, не обладая автономностью):

- «галлюцинируя», давать неверные сведения, касающиеся здоровья человека, законов, событий, технологий, других людей
- вызывать необоснованное доверие и манипулировать человеком
- переубеждать, побуждать человека к действиям, не выгодным ему
- поддерживать предрассудки и лженаучные представления
- поддерживать пропагандистские медиа-кампании
- неконтролируемо влиять на формирование мировоззрения у подростков
- оказывать депрессивное воздействие на психику

Шаги практического решения задач AI/DS/ML

Формализация постановки, «ДНК» задачи

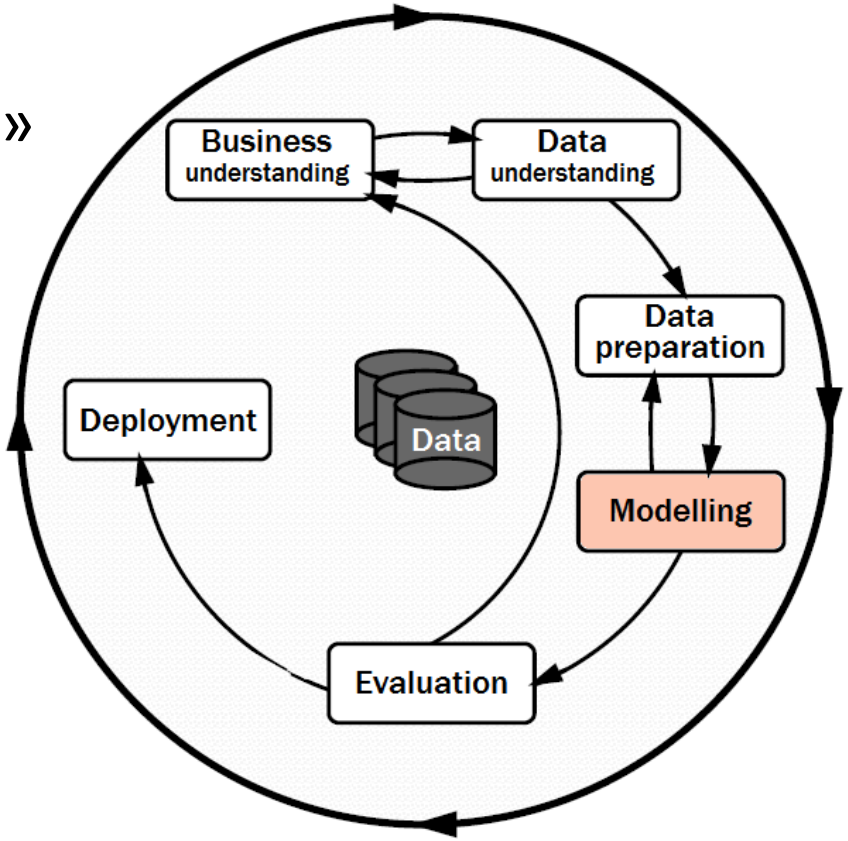
- Дано: выборка «объекты-признаки → ответы»
- Найти: предсказательная модель
- Критерии: качество предсказаний, KPI

Моделирование

- предобработка и векторизация данных
- формализация модели
- оптимизация (обучение) модели
- оценивание и выбор моделей

Внедрение

- оценивание качества оффлайн и онлайн
- интеграция с бизнес-процессами



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

Особенности реальных данных

В реальных приложениях данные бывают …

- разнородные (признаки измерены в разных шкалах)
- неполные (признаки измерены не все, имеются пропуски)
- неточные (признаки измерены с погрешностями)
- противоречивые (объекты одинаковые, ответы разные)
- избыточные (сверхбольшие, не помещаются в память)
- недостаточные (объектов меньше, чем признаков)
- неструктурированные (нет признаковых описаний)
- «грязные» (ошибочные, грубо не соответствующие истине)

со всем этим можно работать ☺

но только не с грязными данными! ☹

Необходимые условия применения ИИ

- **Полнота, чистота, достоверность данных**
  - Автоматизация и цифровизация бизнес-процессов
  - Улучшение качества данных (от «цифрового чучела» к цифровому двойнику)
  - Трудовая и технологическая дисциплина при работе с данными
- **Культура постановки задач**
  - Понимание бизнес-целей и их формализация через измеримые критерии
  - Предметная экспертиза вместо «абстрактной веры во всемогущий ИИ»
  - Готовность пилотировать новые технологии («data-driven» на всех уровнях)
- **Культура анализа данных**
  - Владение средствами визуализации и понимания данных
  - Тщательный анализ ошибок при выборе моделей
  - Умение находить «простые но гениальные» решения

Выводы: что необходимо знать про ИИ

- ИИ = Имитация Интеллекта, не субъект, а набор технологий
- Применение начинается с постановки задачи *Дано-Найти-Критерий*
- Главное — достоверность, полнота, чистота данных
- Новые применения могут требовать новой математики, но такое происходит всё реже, теперь AI/DS/ML — область инженерная
- *Глубокие нейронные сети* нужны в основном для обучаемой векторизации сложно структурированных данных
- *Генеративные модели текста* — не интеллект, а языковой интерфейс к знаниям человечества, избыточным и противоречивым

Рекомендуемые материалы

- *Визильтер Ю. В.* От слабого ИИ к общему универсальному интеллекту (обзор тенденций 2020-2023). Семинар РАИИ и ФИЦ ИУ РАН «Проблемы искусственного интеллекта» 31.01.2024 https://rutube.ru/video/2aad53ec833f19918c1593398a2a1b88/
- Не пропустите открытие тысячелетия! // Vital Math, 13 января 2024, https://www.youtube.com/watch?v=JZjH0it9Jyg
- Report: AI Decrypted: A Guide for Navigating AI Developments in 2024, January 24, 2024 (Навигатор по ИИ-ландшафту от DENTONS GLOBAL ADVISORS) https://www.albrightstonebridge.com/news/report-ai-decrypted-guide-navigating-ai-developments-2024
- 5 идей применения ИИ в вашем бизнесе прямо сейчас, 5 октября 2023. https://dzen.ru/a/ZR6ZeK5B3lL6OxXv
- *Воронцов К. В.* Лекции по машинному обучению. www.MachineLearning.ru, 2004-2023.
- *Гарбук С.В., Губинский А.М.* Искусственный интеллект в ведущих странах мира: стратегии развития и военное применение. Знание, 2020.
- *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.