

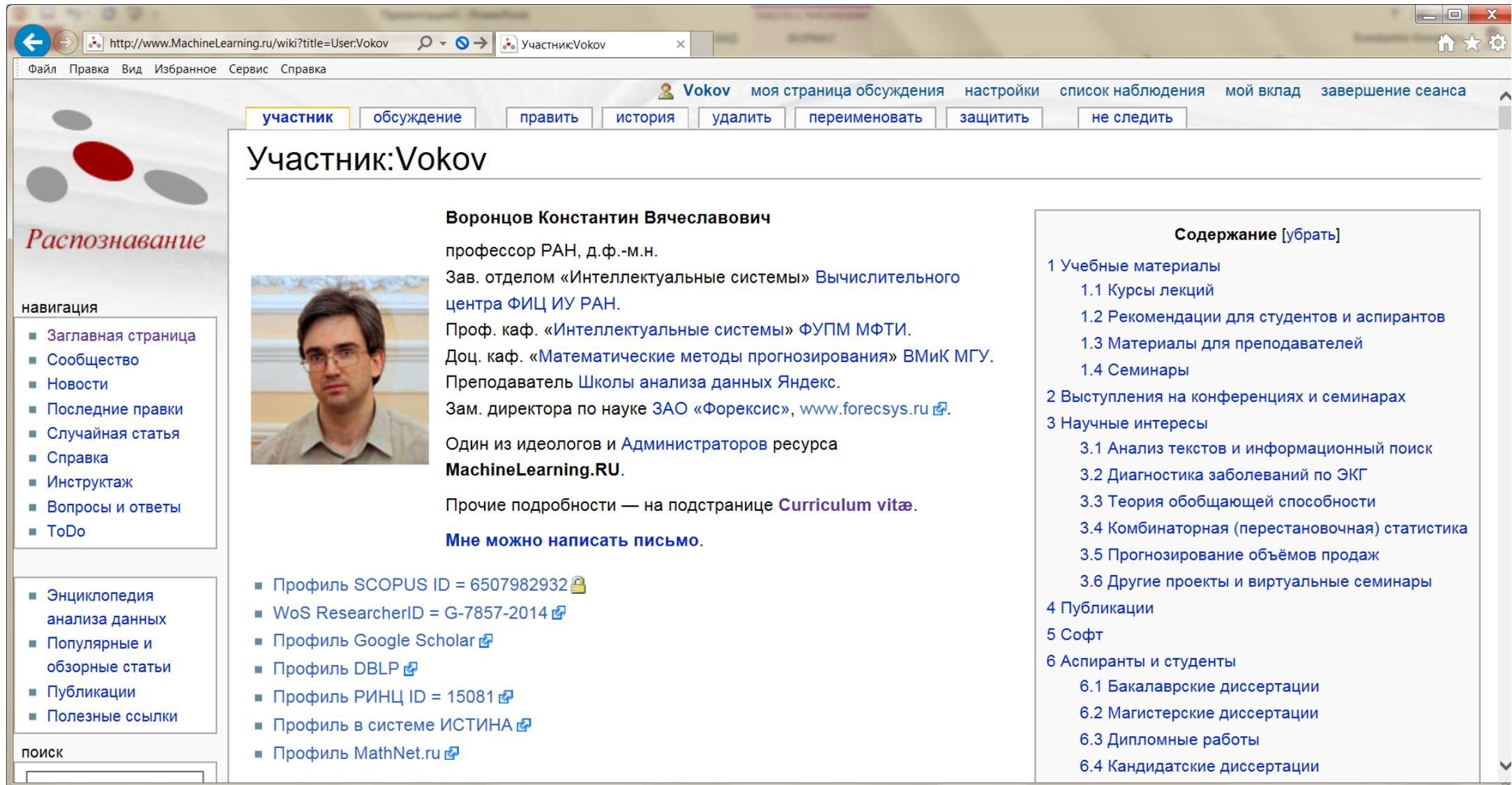
# Машинное обучение: шаг в цифровую экономику

*Воронцов Константин Вячеславович*

- Московский Физико-Технический Институт ●
- Вычислительный Центр им. А.А.Дородницына ФИЦ ИУ РАН ●
  - ШАД Яндекс ●

[voron@forecsys.ru](mailto:voron@forecsys.ru)

# Докладчик: *Воронцов Константин Вячеславович*



The screenshot shows a web browser window displaying the profile page for 'Участник:Vokov' on the Machine Learning RU wiki. The browser's address bar shows the URL 'http://www.MachineLearning.ru/wiki?title=User:Vokov'. The page features a navigation menu at the top with options like 'участник', 'обсуждение', 'править', etc. The main content area includes a profile picture of a man with glasses, a list of his professional affiliations (e.g., professor at RAS, head of the 'Intelligent Systems' department), and a list of external identifiers (SCOPUS ID, WoS ResearcherID, Google Scholar, etc.). A table of contents on the right side lists various sections such as 'Учебные материалы', 'Выступления на конференциях', and 'Публикации'.

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

Участник:Vokov

Файл Правка Вид Избранное Сервис Справка

Vokov моя страница обсуждения настройки список наблюдения мой вклад завершение сеанса

участник обсуждение править история удалить переименовать защитить не следить

## Участник:Vokov

**Воронцов Константин Вячеславович**  
профессор РАН, д.ф.-м.н.  
Зав. отделом «Интеллектуальные системы» Вычислительного центра ФИЦ ИУ РАН.  
Проф. каф. «Интеллектуальные системы» ФУПМ МФТИ.  
Доц. каф. «Математические методы прогнозирования» ВМиК МГУ.  
Преподаватель Школы анализа данных Яндекс.  
Зам. директора по науке ЗАО «Форексис», [www.forecsys.ru](http://www.forecsys.ru).

Один из идеологов и Администраторов ресурса **MachineLearning.RU**.

Прочие подробности — на подстранице **Curriculum vitae**.

[Мне можно написать письмо.](#)

навигация

- Заглавная страница
- Сообщество
- Новости
- Последние правки
- Случайная статья
- Справка
- Инструктаж
- Вопросы и ответы
- ToDo

- Энциклопедия анализа данных
- Популярные и обзорные статьи
- Публикации
- Полезные ссылки

поиск

### Содержание [убрать]

- 1 Учебные материалы
  - 1.1 Курсы лекций
  - 1.2 Рекомендации для студентов и аспирантов
  - 1.3 Материалы для преподавателей
  - 1.4 Семинары
- 2 Выступления на конференциях и семинарах
- 3 Научные интересы
  - 3.1 Анализ текстов и информационный поиск
  - 3.2 Диагностика заболеваний по ЭКГ
  - 3.3 Теория обобщающей способности
  - 3.4 Комбинаторная (перестановочная) статистика
  - 3.5 Прогнозирование объёмов продаж
  - 3.6 Другие проекты и виртуальные семинары
- 4 Публикации
- 5 Софт
- 6 Аспиранты и студенты
  - 6.1 Бакалаврские диссертации
  - 6.2 Магистерские диссертации
  - 6.3 Дипломные работы
  - 6.4 Кандидатские диссертации

# Машинное обучение: шаг в цифровую экономику

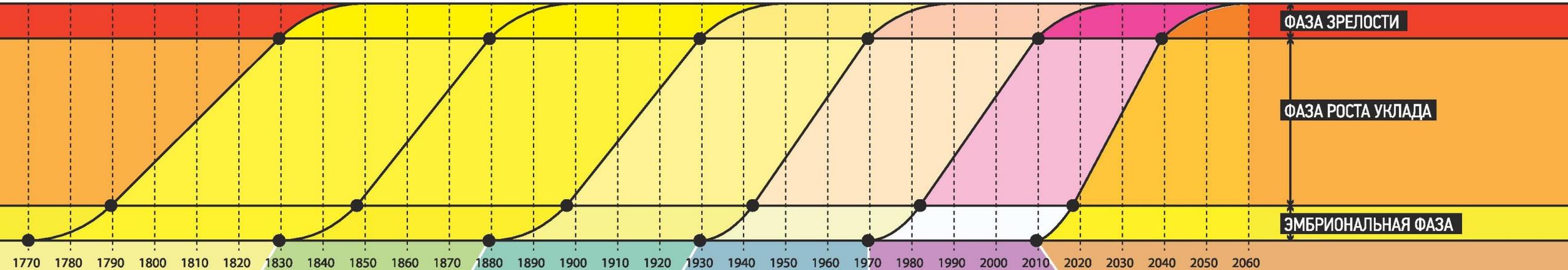
1. Знакомство с машинным обучением
  - Технология будущего или очередной хайп?
  - Основные понятия
  - Примеры прикладных задач
2. Как оно менялось само
3. Как оно теперь меняет мир
4. Как мы решаем задачи

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

Клаус Мартин Шваб,  
президент Всемирного  
экономического форума



# Технологические уклады и революции



## ПЕРВЫЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

**Основной ресурс:** энергия воды

**Главная отрасль:** текстильная промышленность

**Ключевой фактор:** текстильные машины

**Достижение уклада:** механизация фабричного производства

## ВТОРОЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

**Основной ресурс:** энергия пара, уголь

**Главная отрасль:** транспорт, чёрная металлургия

**Ключевой фактор:** паровой двигатель, паровые приводы станков

**Достижения уклада:** рост масштабов производства, развитие транспорта

**Гуманитарное преимущество:** постепенное освобождение человека от тяжёлого ручного труда

## ТРЕТИЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

**Основной ресурс:** электрическая энергия

**Главная отрасль:** тяжелое машиностроение, электротехническая промышленность

**Ключевой фактор:** электродвигатель

**Достижения уклада:** концентрация банковского и финансового капитала; появление радиосвязи, телеграфа; стандартизация производства;

**Гуманитарное преимущество:** повышение качества жизни

## ЧЕТВЕРТЫЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

**Основной ресурс:** энергия углеводородов, начало ядерной энергетики

**Основные отрасли:** автомобилестроение, цветная металлургия, нефтепереработка, синтетические полимерные материалы

**Ключевой фактор:** двигатель внутреннего сгорания, нефтехимия

**Достижения уклада:** массовое и серийное производство

**Гуманитарное преимущество:** развитие связи, транснациональных отношений, рост производства продуктов народного потребления

## ПЯТЫЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

**Основной ресурс:** атомная энергетика

**Основные отрасли:** электроника и микроэлектроника, информационные технологии, программное обеспечение, телекоммуникации, освоение космического пространства

**Ключевой фактор:** микроэлектронные компоненты

**Достижения уклада:** индивидуализация производства и потребления

**Гуманитарное преимущество:** глобализация, скорость связи и перемещения

## ШЕСТОЙ ТЕХНОЛОГИЧЕСКИЙ УКЛАД

Все составляющие нового технологического уклада носят характер прогноза

**Основные отрасли:** нано- и биотехнологии, наноэнергетика, молекулярная, клеточная и ядерная технологии, нанобиотехнологии, биомиметика, нанобионика, нанотроника и другие наноразмерные производства; новые медицина, бытовая техника, виды транспорта и коммуникаций, использование стволовых клеток, инженерия живых тканей и органов, восстановительная хирургия и медицина

**Ключевой фактор:** микроэлектронные компоненты

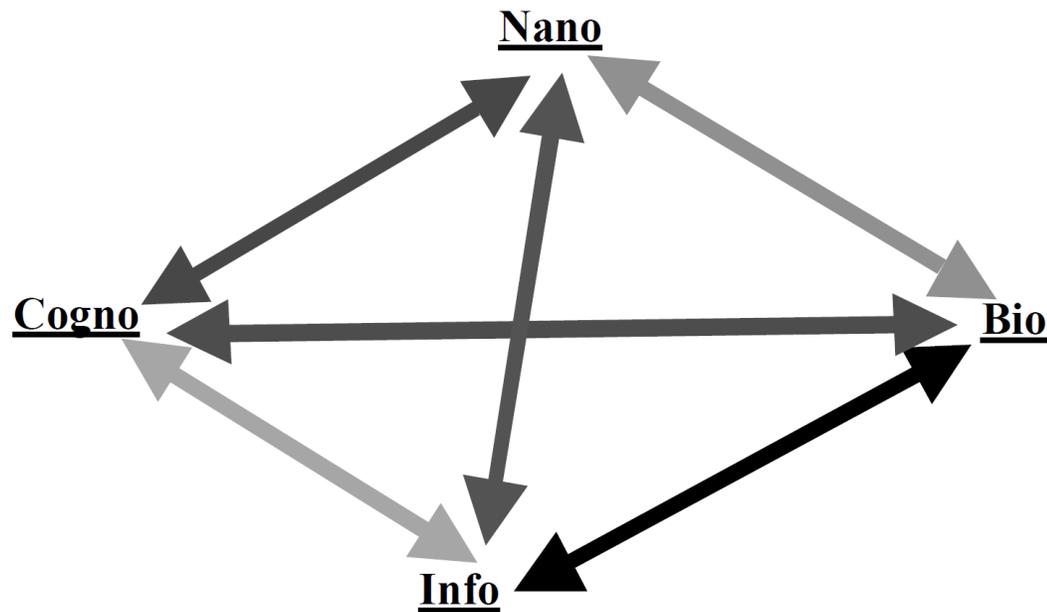
**Достижения уклада:** индивидуализация производства и потре-

бления, резкое снижение энергоёмкости и материалоемкости производства, конструирование материалов и организмов с заранее заданными свойствами

**Гуманитарное преимущество:** существенное увеличение продолжительности и качества жизни человека и животных

**На 2010 год** доля производительных сил пятого технологического уклада в наиболее развитых странах составляла примерно 60%, четвертого — 20%, шестого — около 5%. По последним расчетам учёных, шестой технологический уклад в этих странах фактически наступит в 2014–2018 годах.

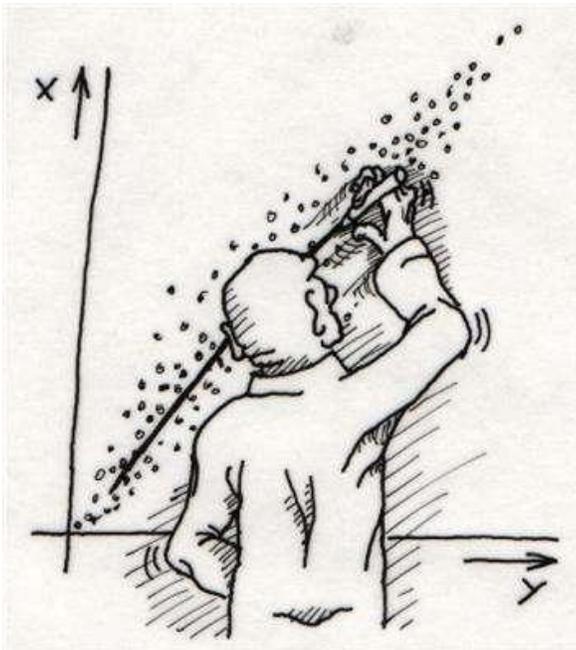
# Конвергенция технологий шестого уклада



- Info – технологии 5-го уклада
- Опережающее развитие Info
- Основные элементы Info
  - Интернет
  - Искусственный интеллект
  - Машинное обучение
  - Большие данные
  - Высокопроизводительные вычисления
  - Робототехника

# Машинное обучение – это ...

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



- *проведение функции через заданные точки в сложно устроенных пространствах*
- математическое моделирование, когда данных много, знаний мало
- тысячи алгоритмов
- около 100 000 научных публикаций в год

# Основная задача машинного обучения

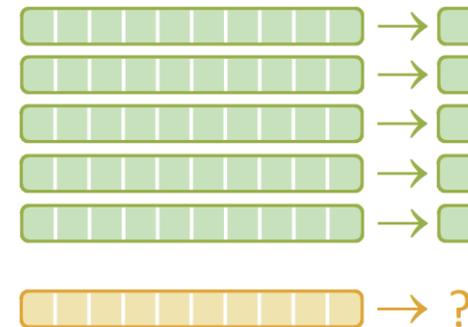
## Этап №1 – обучение с учителем

- **На входе:**  
данные – выборка прецедентов «объект → ответ»
- **На выходе:**  
алгоритм, по любому объекту предсказывающий ответ

*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 / лечение / риск исхода



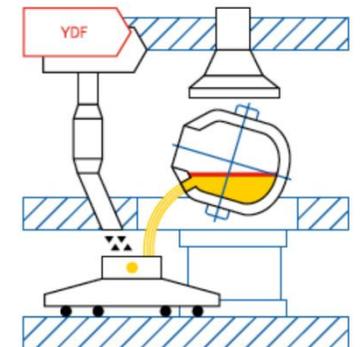
- **Поиск месторождений полезных ископаемых:**

объект – данные о геологии района  
ответ – есть/нет месторождение



- **Управление технологическими процессами:**

объект – данные о сырье и управляющих параметрах  
ответ – количество/качество полезного продукта



# Примеры задач машинного обучения

- **Кредитный скоринг:**

объект – данные о заёмщике

ответ – вероятность дефолта, решение по кредиту



- **Предсказание оттока клиентов:**

объект – данные о клиенте на момент времени  $t$

ответ – уйдёт ли клиент к моменту времени  $t + \Delta$



- **Прогнозирование объёмов продаж:**

объект – данные о продажах на момент времени  $t$

ответ – объём спроса в интервале от  $t$  до  $t + \Delta$



# Примеры задач машинного обучения

- **Информационный поиск в Интернете:**

объект – данные о паре «запрос и документ»

ответ – оценка релевантности документа запросу



- **Продажа рекламы в Интернете:**

объект – данные о тройке «пользователь, страница, баннер»

ответ – оценка вероятности клика

- **Рекомендательные системы в Интернете / TV:**

объект – данные о паре «пользователь, товар / фильм»

ответ – оценка вероятности покупки / просмотра



# Примеры задач машинного обучения

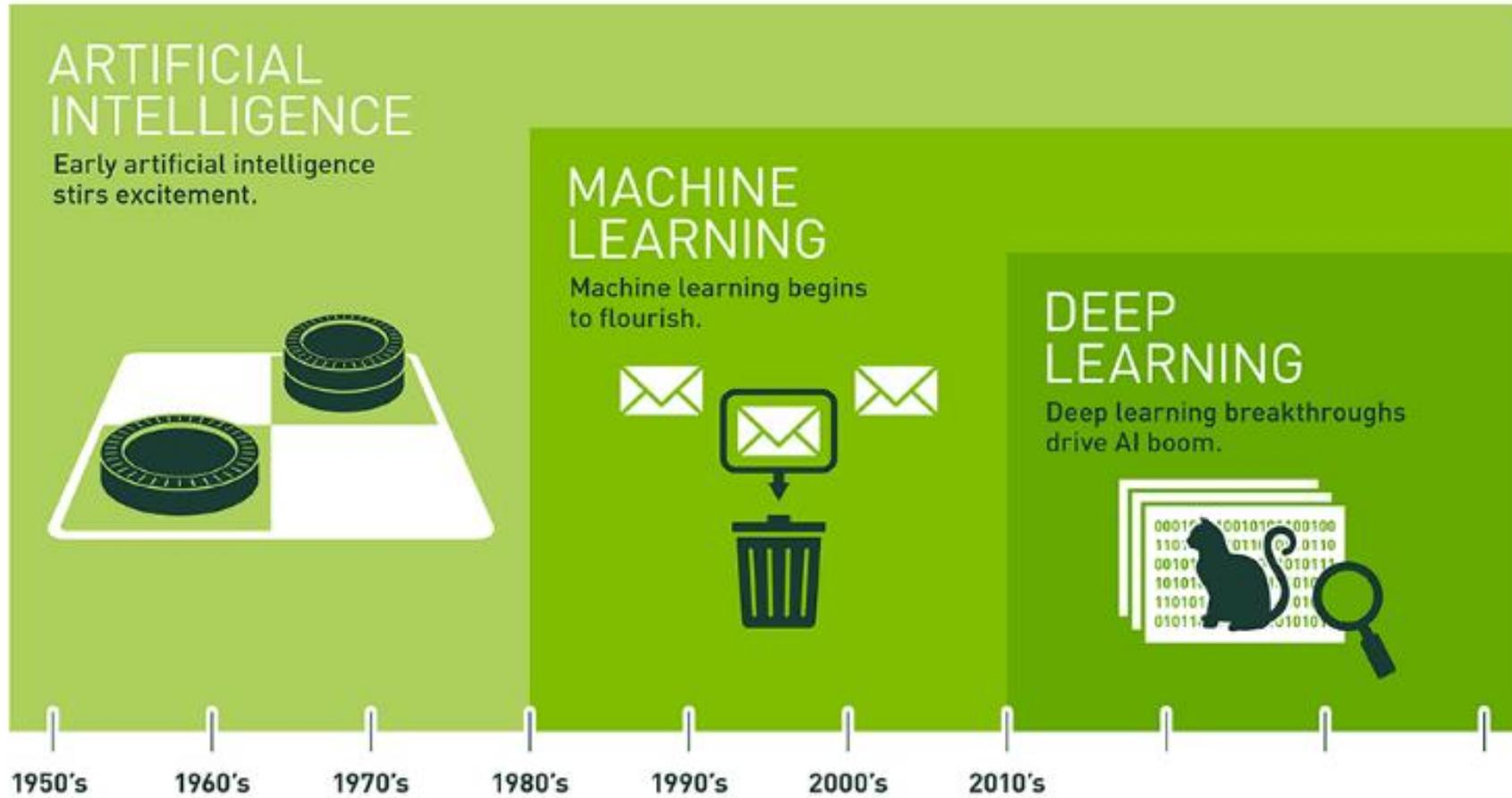
- **Статистический машинный перевод:**  
объект – предложение на естественном языке  
ответ – его перевод на другой язык
- **Перевод речи в текст:**  
объект – аудиозапись речи человека  
ответ – текстовая запись речи
- **Компьютерное зрение:**  
объект – изображение предмета в видеопоследовательности  
ответ – решение (объехать, остановиться, игнорировать)

*Прогресс в этих  
областях связан с  
«Большими данными»,  
англ. «Big Data»*

# Машинное обучение: шаг в цифровую экономику

1. Знакомство с машинным обучением
2. Как оно менялось само
  - Эволюция идей и направлений ИИ
  - Бум искусственного интеллекта 201X
  - Три предпосылки этого бума
3. Как оно теперь меняет мир
4. Как мы решаем задачи

# Эволюция искусственного интеллекта



*Глубокое обучение  
– одна из новейших  
технологий  
машинного  
обучения*

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

# Глубокие нейронные сети обеспечили прорыв в компьютерном зрении

- Распознавание кадров с котами на видео из Youtube
- Автономные летательные аппараты



# Машинное обучение, большие данные «и много других страшных слов»

- Статистический анализ данных (Statistical Data Analysis)
- Искусственный интеллект (Artificial Intelligence) 1955
- Распознавание образов (Pattern Recognition)
- Машинное обучение (Machine Learning) 1959
- Статистическое обучение (Statistical Learning)
- Интеллектуальный анализ данных (Data Mining) 1989
- Бизнес-аналитика (Business Intelligence, Business Analytics)
- Предсказательная аналитика (Predictive Analytics) 2007
- Большие данные (Big Data) 2008
- Аналитика больших данных (Big Data Analytics)
- Наука о данных (Data Science) 2011

# Бум искусственного интеллекта

**1997:** IBM Deep Blue обыграл чемпиона мира по шахматам

**2005:** Беспилотный автомобиль: DARPA Grand Challenge

**2006:** Google Translate – статистический машинный перевод

**2011:** 40 лет DARPA CALO привели к созданию Apple Siri

**2011:** IBM Watson победил в ТВ-игре «Jeopardy!»

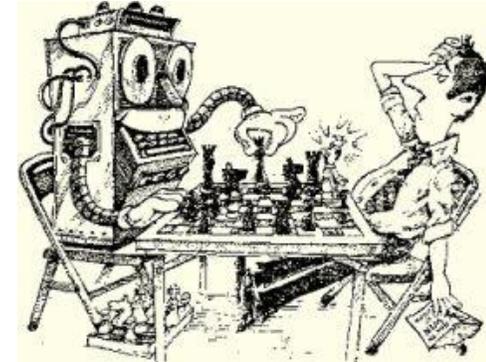
**2011–2015:** ImageNet: 25% → 3,5% ошибок против 5% у людей

**2015:** Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана

**2016:** DeepMind, OpenAI: динамическое обучение играм Atari

**2016:** Google DeepMind обыграл чемпиона мира по игре го

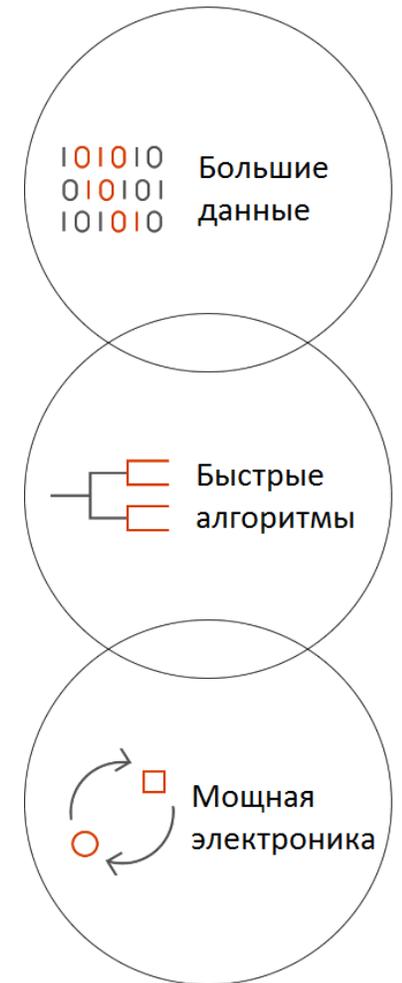
**2017:** OpenAI обыграл чемпиона мира по компьютерной игре Dota 2



# Три предпосылки этого бума

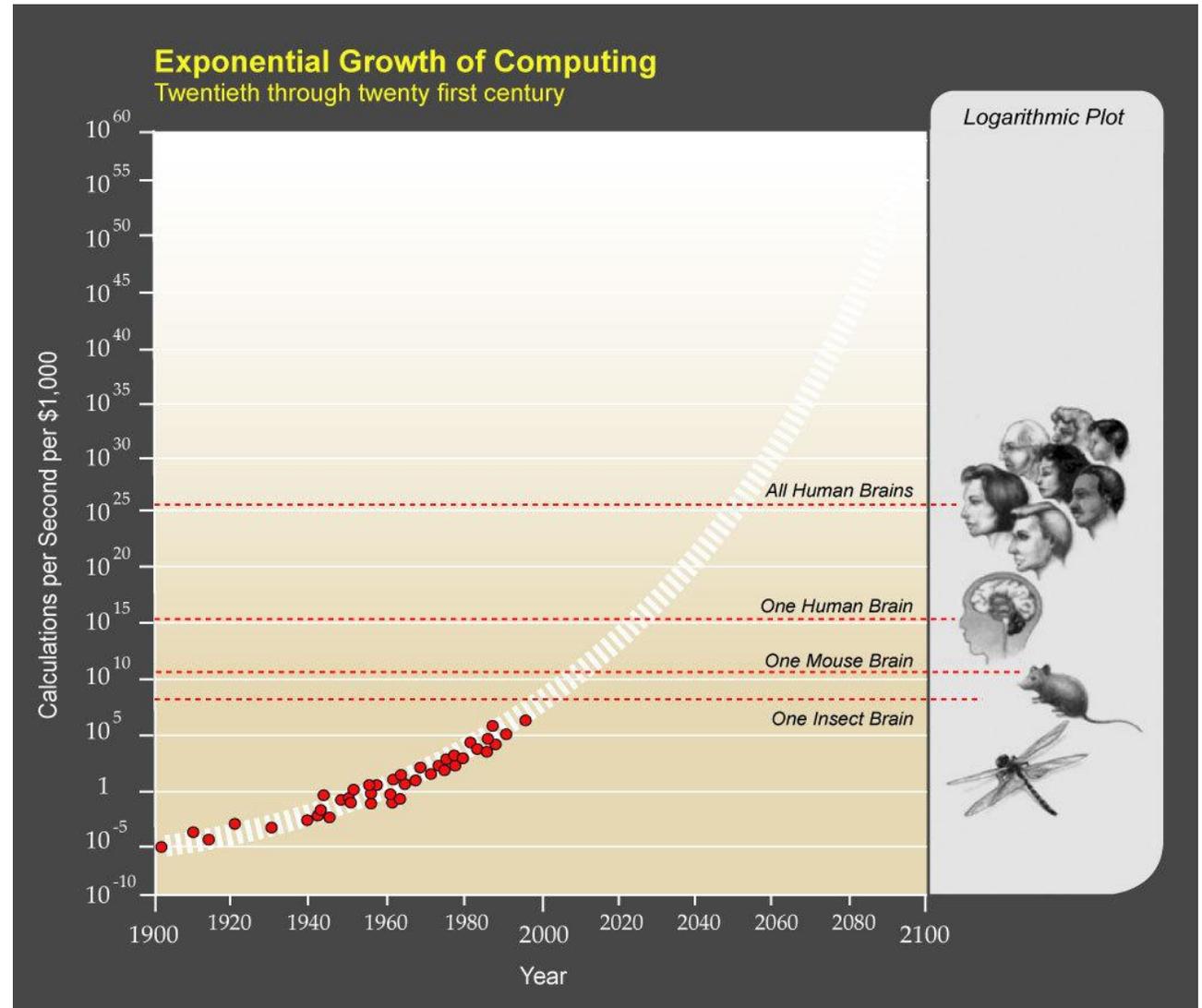
– три перехода количества в качество:

- Повсеместное применение компьютерных технологий  
→ *Накопление больших выборок данных*
- Развитие математических методов и алгоритмов  
→ *Накопление критической массы опыта*
- Достижения микроэлектроники  
→ *Рост вычислительных мощностей по закону Мура*



# Закон Мура

Закон  
ускоряющейся  
отдачи  
(Рэймонд Курцвейл)



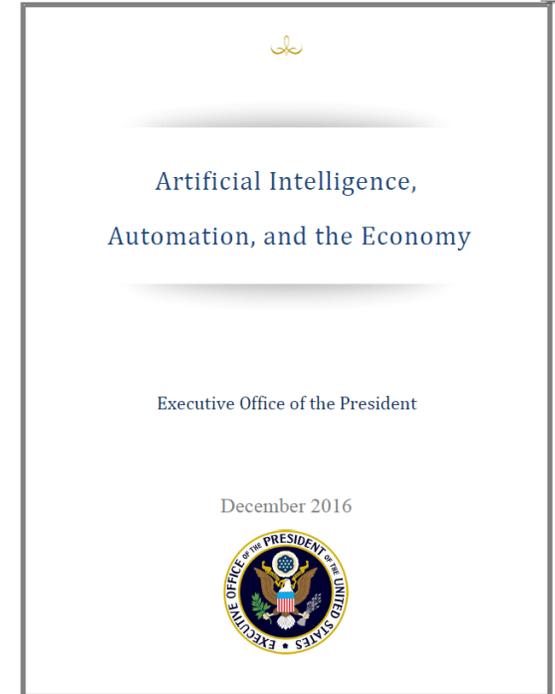
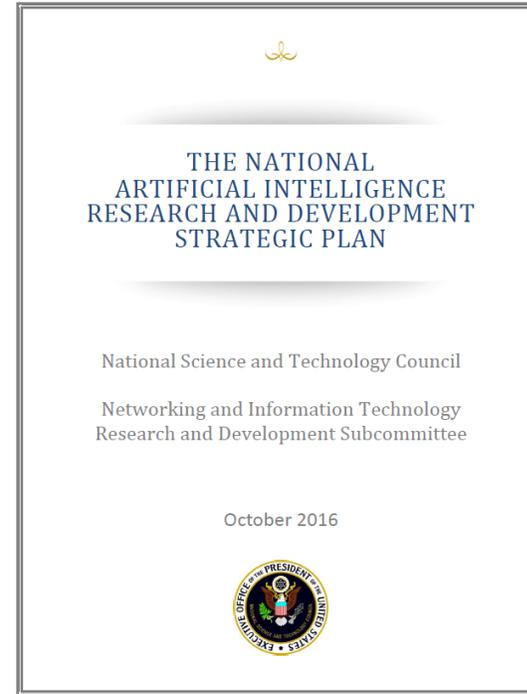
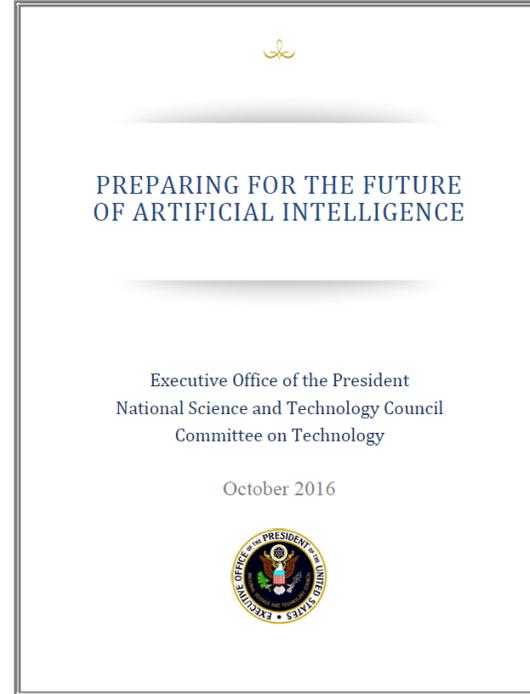
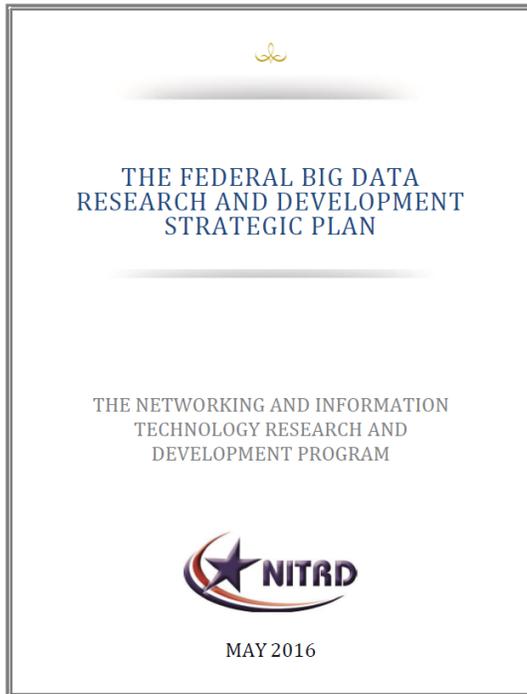
# Машинное обучение: шаг в цифровую экономику

1. Знакомство с машинным обучением
2. Как оно менялось само
3. Как оно теперь меняет мир
  - Отчёты Белого Дома США конца 2016 г.
  - Открытые данные
  - Открытый код
4. Как мы решаем задачи

# Технологии ИИ, которые меняют мир



# Отчёты Белого дома США, май-октябрь 2016



## Основные выгоды ИИ

- Автоматизация и сокращение издержек повсеместно
- Автономный транспорт и роботизация
- Оптимизация логистики и цепей поставок
- Оптимизация энергетических и транспортных сетей
- Сенсорные сети, мониторинг сельского хозяйства
- Автоматизация банковских услуг и посреднической деятельности
- Информационные сервисы и распределённая экономика
- Персональная медицина, улучшение клинических практик
- Персональные образовательные траектории, социальная инженерия
- Автономные системы вооружений

## Государственное финансирование ИИ

Правительство США выделяет на исследования по ИИ

- \$1.1 млрд. в 2015
- \$1.2 млрд. в 2016

(без учёта вложений в ИИ со стороны частных компаний США).

Согласно отчёту, для США необходимо в 2-4 раза больше.

\$1 млрд. выделяет Южная Корея,

\$1 млрд. выделяет компания Toyota,

\$1 млрд. инвестиций привлекла компания Илона Маска OpenAI.

## Потенциальные опасности ИИ

- ИИ превзойдёт человеческий (прогнозы 2035-2050 г.)
  - возможно... но ведь машины давно сильнее и быстрее людей
- Власть захватят машины (или плохие парни с машинами)
  - нет: открытый код + ограничение автономности ИИ + аскетичность ИИ
- Нас завоюют другие страны с более мощным ИИ
  - возможно... если мы не будем развивать свои технологии
- Машины лишат людей рабочих мест
  - нет, ИИ одновременно создаёт новые профессии и рабочие места  
кроме того, можно уменьшать продолжительность рабочего дня
- Увеличится социальное неравенство
  - нет, если своевременно проводить массовую переподготовку кадров

## Некоторые из 23 рекомендаций

- #1. Организации должны активно развивать партнёрство с научными коллективами для эффективного использования данных.
- #2. В приоритетном порядке развивать стандарты открытых данных для привлечения научного сообщества к решению задач.
- #8. Инвестировать в разработку систем автоматического управления воздушным трафиком.
- #11. Вести постоянный мониторинг развития ИИ в других странах.
- #13. Приоритетно поддерживать фундаментальные и долгосрочные исследования в области искусственного интеллекта.
- #20, #21. Развивать международную кооперацию по ИИ.

## 7 стратегий R&D в области ИИ

1. Долгосрочные инвестиции в исследования в области ИИ
2. Разработка эффективных человеко-машинных систем ИИ
3. Исследование этических, юридических и социальных аспектов ИИ
4. Обеспечение безопасности, надёжности и доверия к системам ИИ
5. Развитие открытых данных и средств разработки ИИ
6. Развитие стандартов и платформ для тестирования ИИ
7. Подготовка квалифицированных кадров в области ИИ

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

# Открытые данные для ИИ

- **Выгоды открытых данных**
  - вовлечение экспертов в решение задач, поиск кадров
  - обучение студентов на востребованных прикладных задачах
  - популяризация научных знаний в смежных областях
- **Конкурсы анализа данных**
  - [www.NetflixPrize.com](http://www.NetflixPrize.com) (2006-2009) – первый крупный конкурс, \$1 млн.
  - [www.kaggle.com](http://www.kaggle.com) – самая известная платформа
  - [www.FakeNewsChallenge.org](http://www.FakeNewsChallenge.org) – один из последних конкурсов
- [DataRing.ru](http://DataRing.ru) – отечественная конкурсная платформа
  - консалтинг по подготовке данных и условий конкурса
  - очистка, отбор, агрегирование, деперсонификация данных

# Открытые данные в России

- Портал «Открытые данные России» <http://data.gov.ru>
- Конкурс «Открытые данные РФ» <http://opendatacontest.ru>
- Росстат <http://www.gks.ru/opendata>
- Центральный Банк РФ <http://www.cbr.ru/statistics>
- Министерство финансов РФ <http://minfin.ru/opendata>
- Министерство транспорта РФ <http://www.mintrans.ru/opendata>
- Министерство образования и науки:  
<http://открытые-данные.минобрнауки.рф/opendata>

**Подумать:** какие задачи можно решать на этих данных?

# Открытое программное обеспечение ИИ

- Выгоды открытого кода
  - Снижение издержек, ускорение разработки и внедрения
  - Координация усилий исследователей и разработчиков
  - Снижение технологических барьеров для выхода на рынок
- Примеры открытых экосистем машинного обучения
  - Python + SciPy + Sckit-learn
  - Java + WEKA + RapidMiner
  - Deductor, Loginom – новая платформа от [BaseGroup.ru](http://BaseGroup.ru)
- Примеры открытых инструментов для нейронных сетей:
  - DeepMind, OpenAI, TensorFlow, Keras, Teano, Torch, Bonsai, ...

# Машинное обучение: шаг в цифровую экономику

1. Знакомство с машинным обучением
2. Как оно менялось само
3. Как оно теперь меняет мир
4. Как мы решаем задачи
  - Особенности реальных данных
  - Особенности постановок задач
  - Этапы решения в стандарте CRISP-DM
  - Проблематика обучаемости
  - Типология задач машинного обучения

# Особенности данных и постановок задач

## Свойства реальных данных:

- 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)
- неструктурированные (нет признаков описаний)
- ~~«грязные» (грубо не соответствующие истине)~~

*Самая  
большая беда  
– «грязные»  
данные*

# Особенности данных и постановок задач

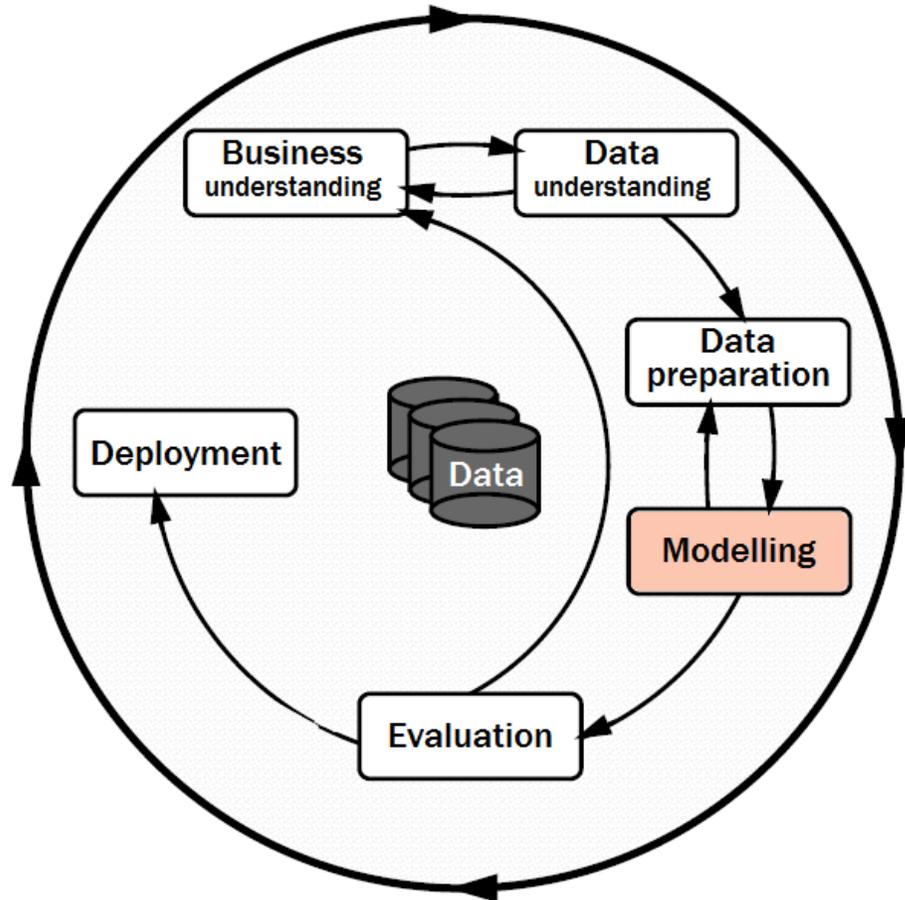
## Идеальный заказчик:

- знает точно, чего хочет
- имеет численные критерии качества (KPI)
- заботится о качестве своих данных
- готов пилотировать новые технологии
- понимает ограничения готовых методов, и что чудес не бывает
- видит уникальность задачи, когда нужна новая математика

*обучение машин  
заставляет  
думать людей*

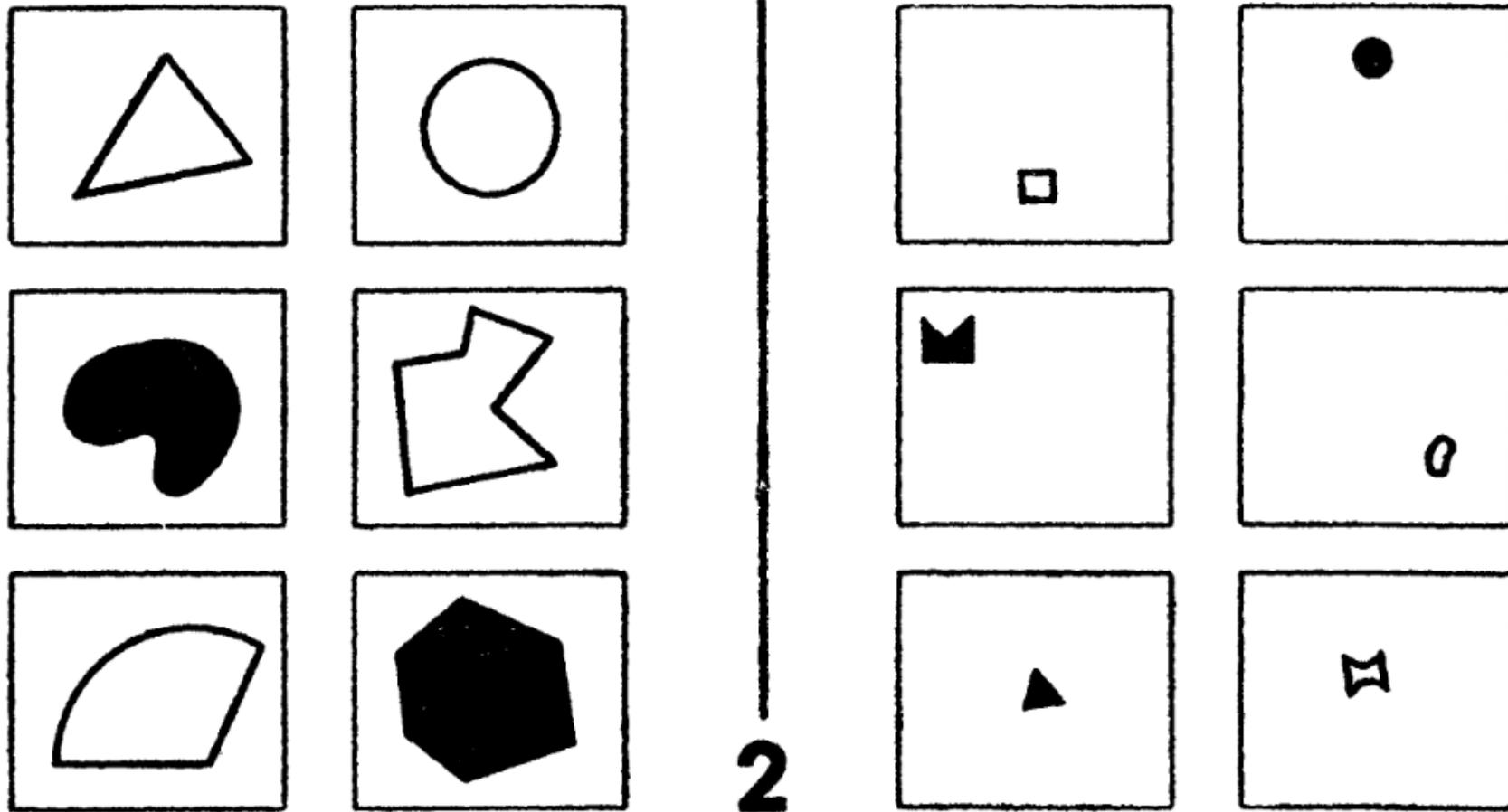
# Этапы решения задач анализа данных

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



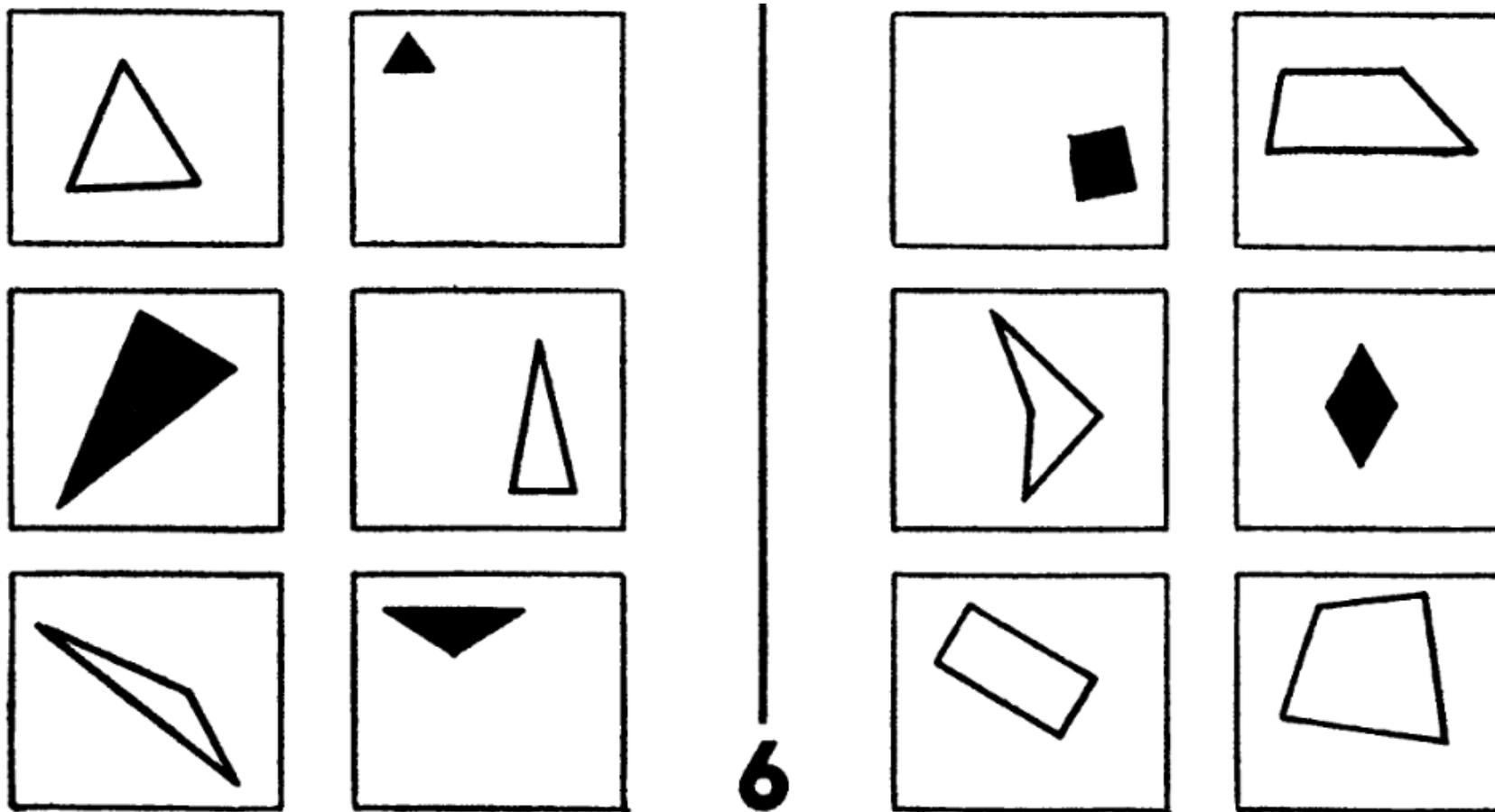
- понимание бизнес-задач
- понимание данных
- предобработка данных
- инженерия признаков
- построение моделей
- оптимизация параметров
- контроль переобучения
- оценивание качества решения
- внедрение и эксплуатация

Тесты Бонгарда (1967). Требуется найти правило классификации.  
Обучающая выборка: по 6 объектов каждого из двух классов.

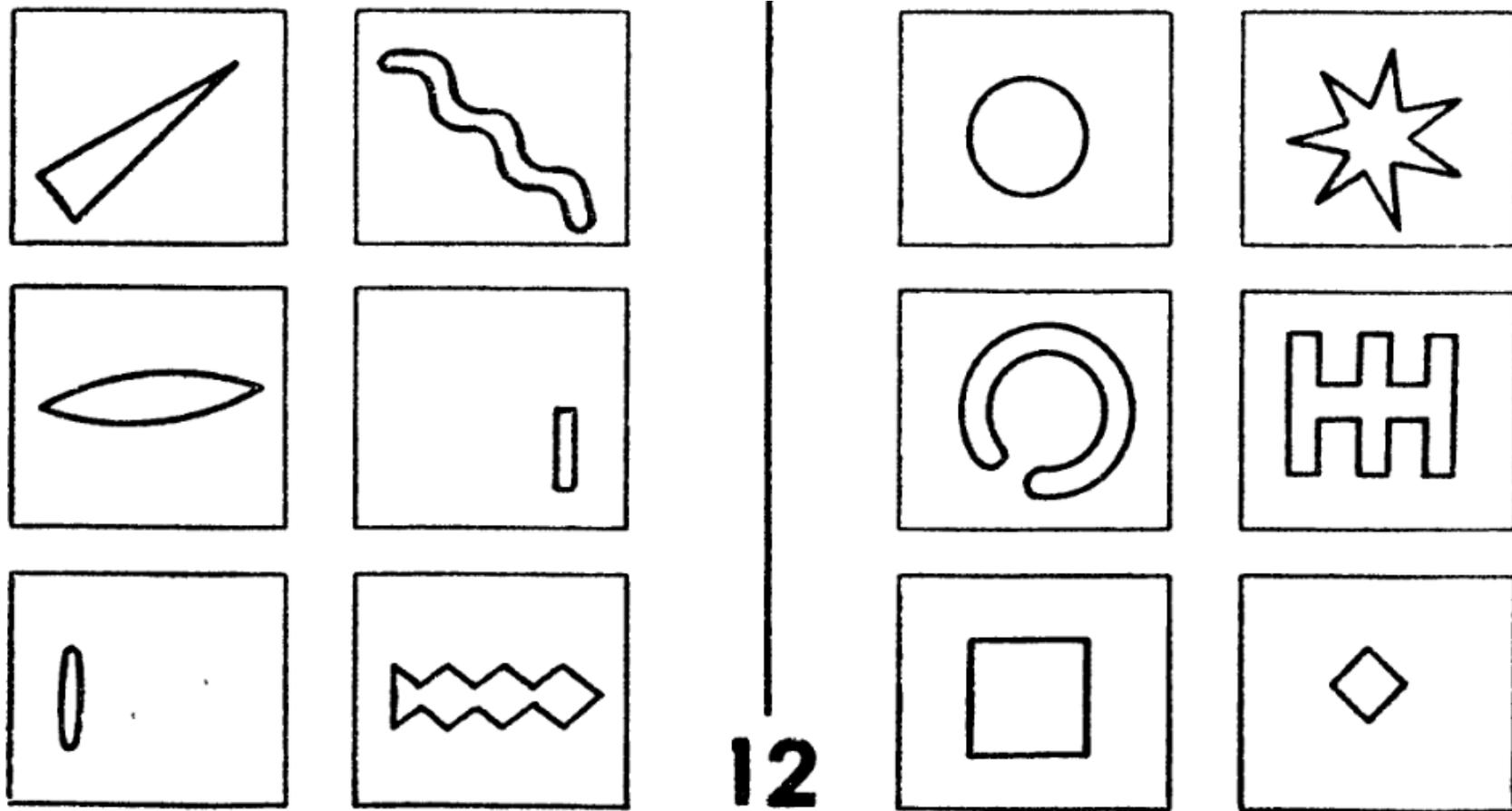


Что даёт нам уверенность, что мы нашли верное правило?

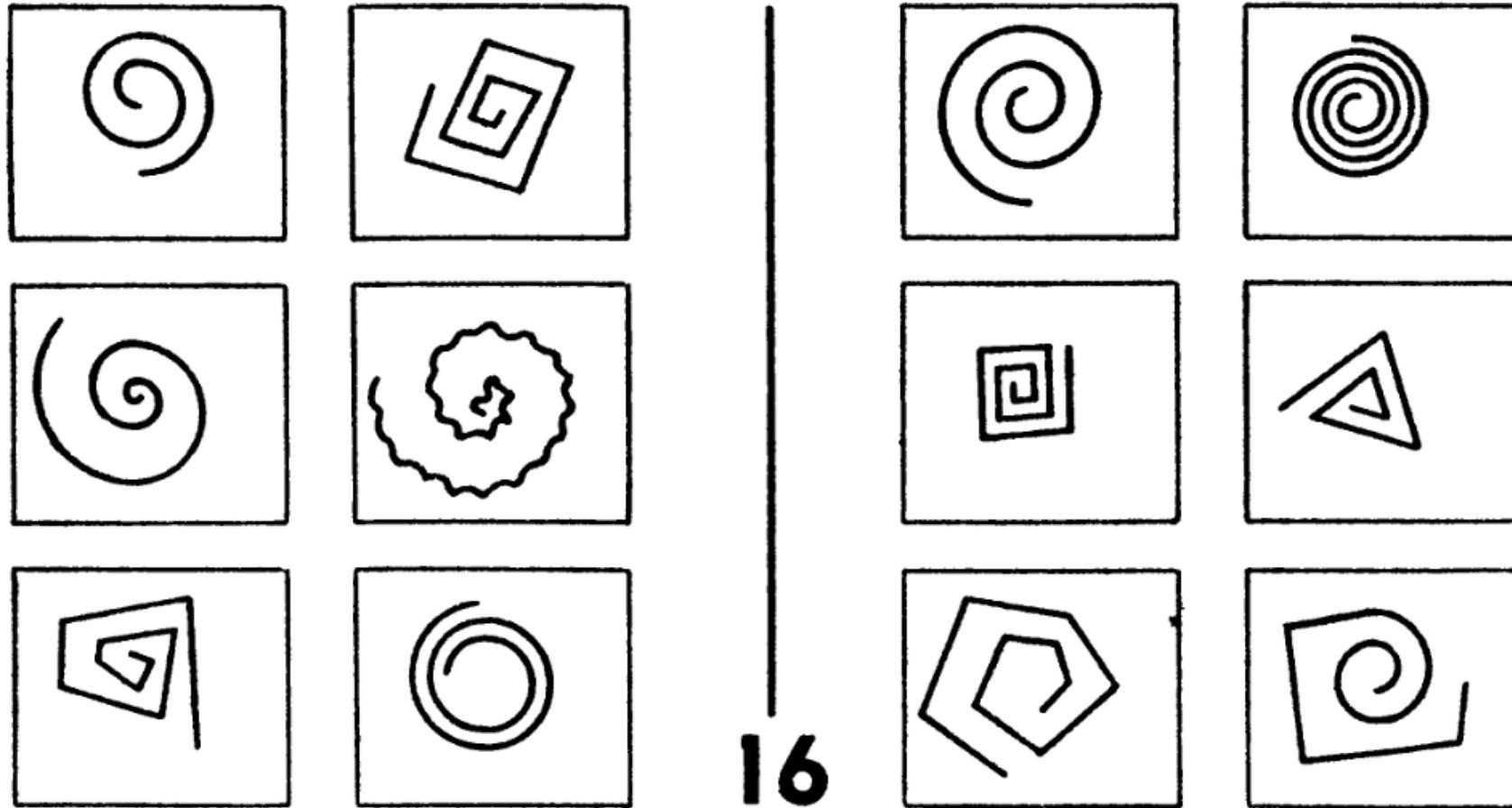
1. Безошибочная классификация примеров обучающей выборки



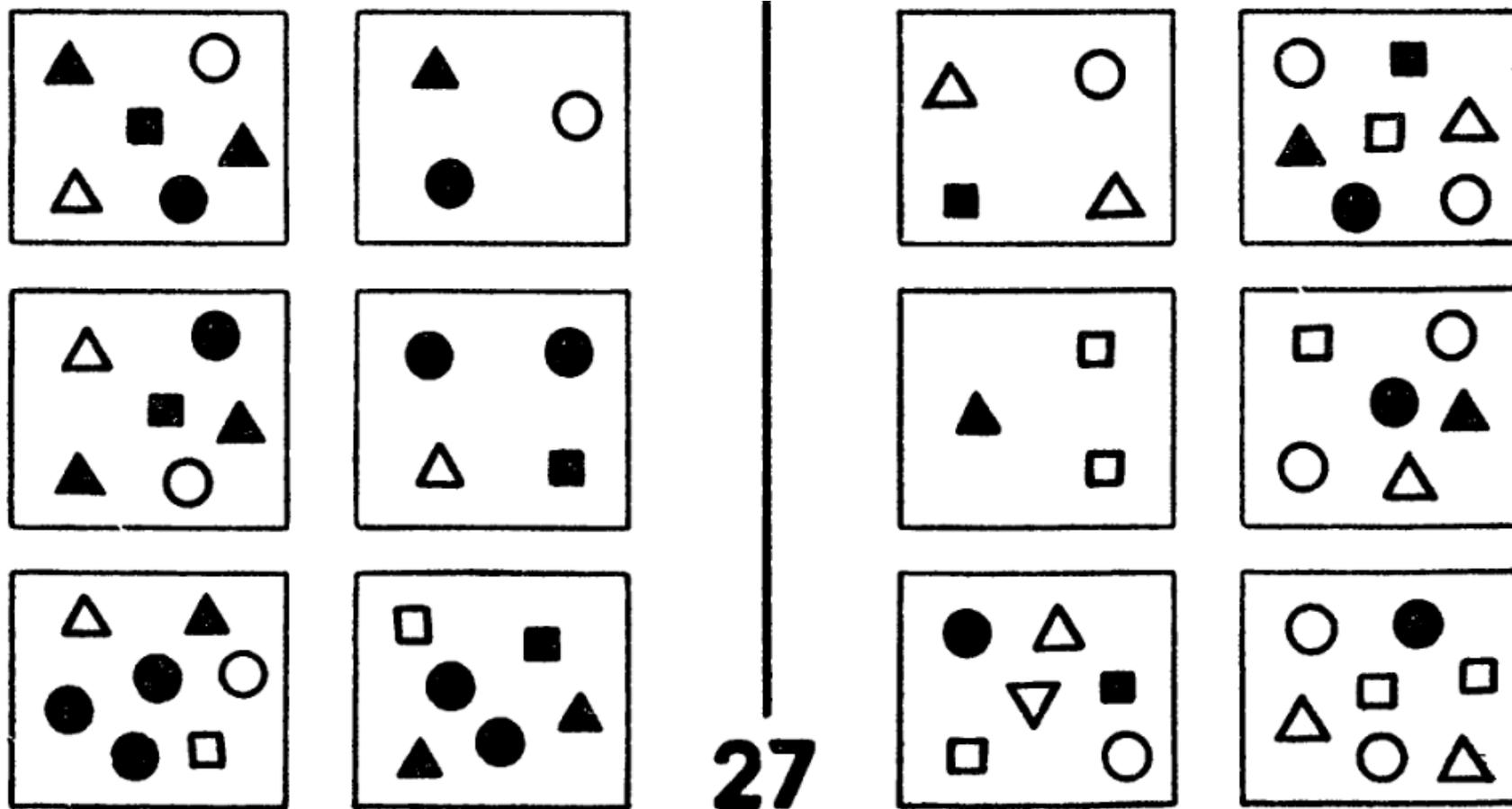
Что ещё даёт нам уверенность, что мы нашли верное правило?  
2. Простота и определённое «изящество» найденного правила.



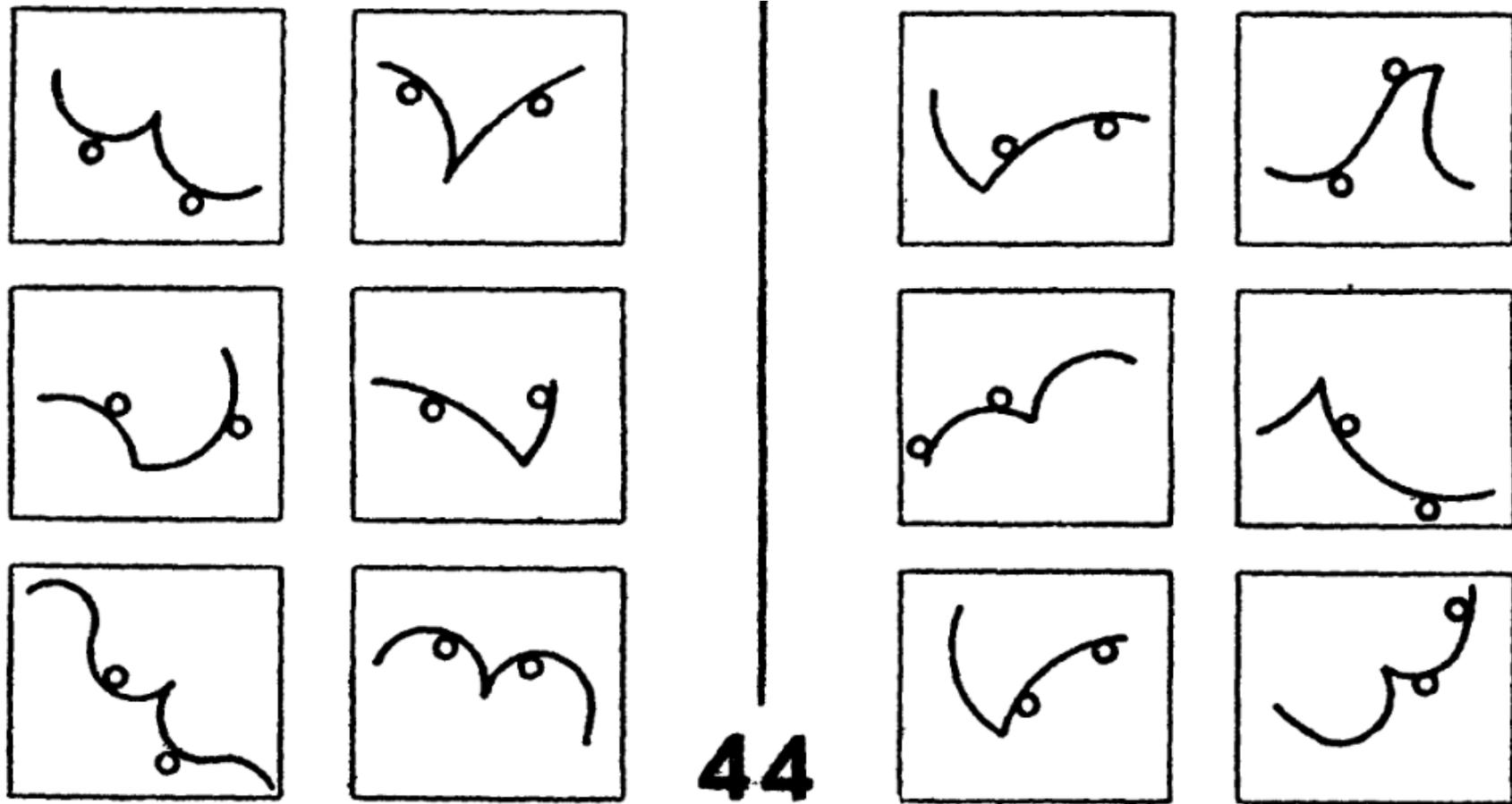
Мы решаем эти задачи почти мгновенно. Чем мы пользуемся?  
Почему для компьютера они столь сложны?



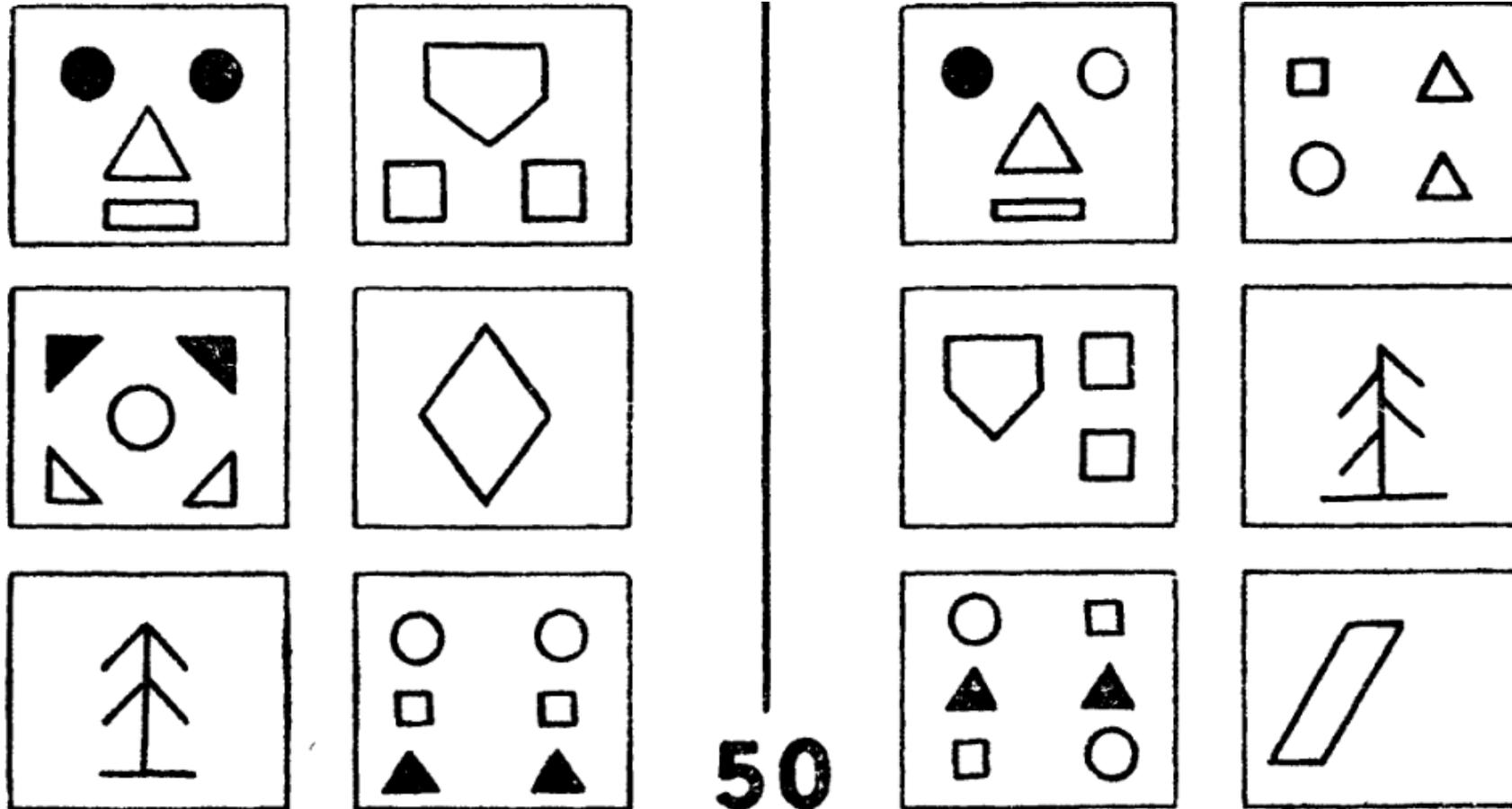
Нужно ли закладывать знания геометрии в явном виде?  
Или возможно выучить геометрические понятия на примерах?



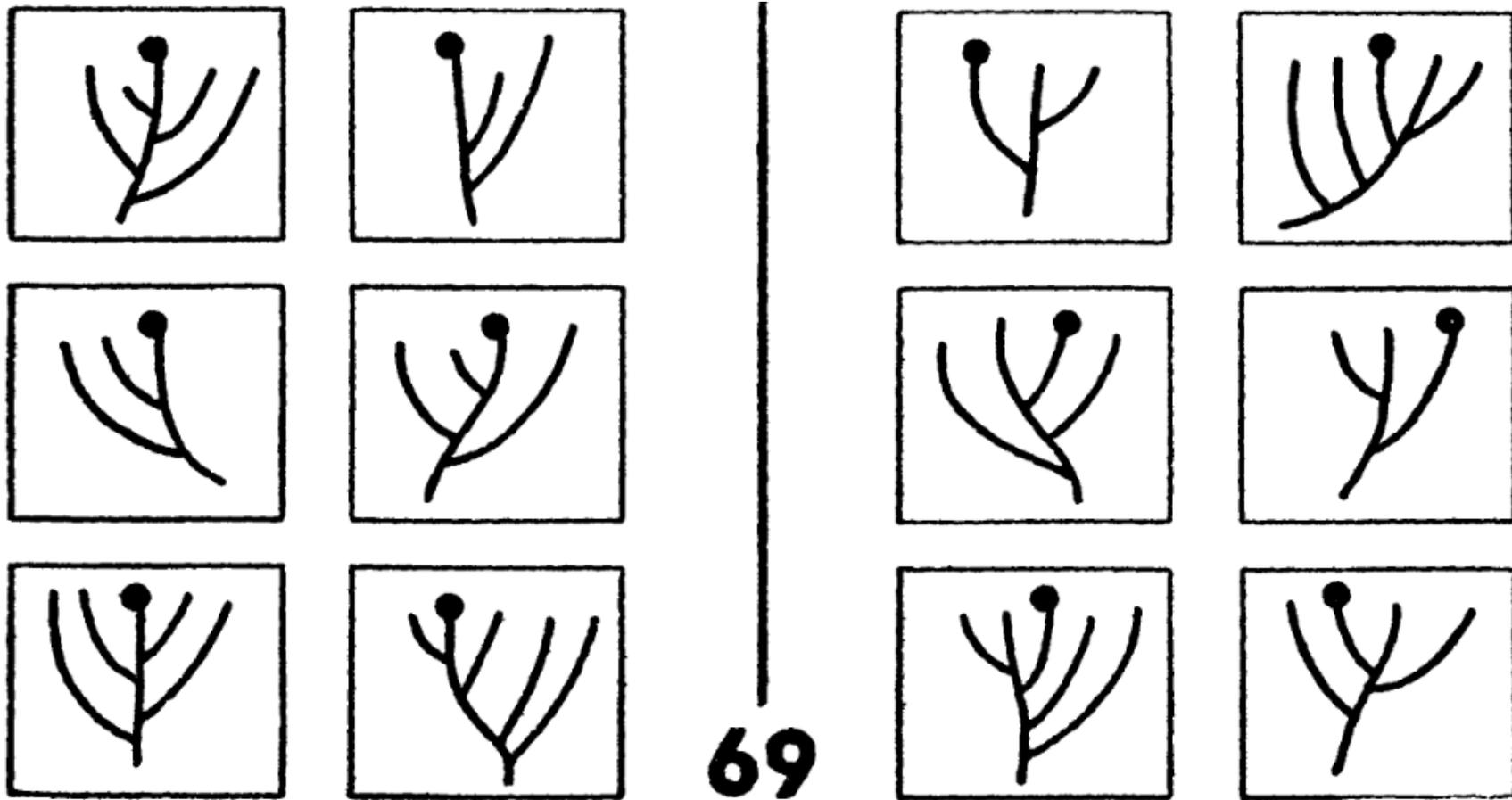
Как вычислять полезные признаки по сложным сырым данным?  
Возможно ли поручить перебор признаков и моделей машине?



Каков риск выбрать по данным неверное правило, *предвзвешенность*?  
Как этот риск зависит от числа примеров и сложности правил?

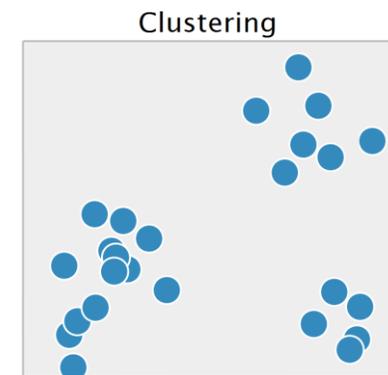
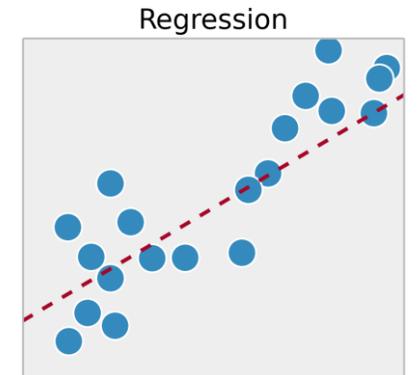
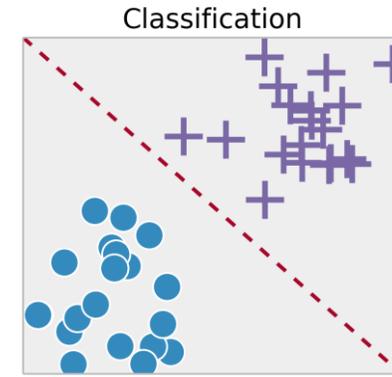


Эти вопросы составляют основу машинного обучения сегодня.  
М.М.Бонгард поставил все эти проблемы в середине 60-х!



# Типология задач машинного обучения

- Обучение с учителем (supervised learning)
  - классификация (classification)
  - регрессия (regression)
  - ранжирование (learning to rank)
  - прогнозирование (forecasting)
- Обучение без учителя (unsupervised learning)
  - кластеризация (clustering)
  - поиск ассоциативных правил (association rule learning)
  - восстановление плотности (density estimation)
  - одноклассовая классификация (anomaly detection)
- Частичное обучение (semi-supervised learning)
  - Обучение с положительными примерами (PU-learning)



# Типология задач машинного обучения

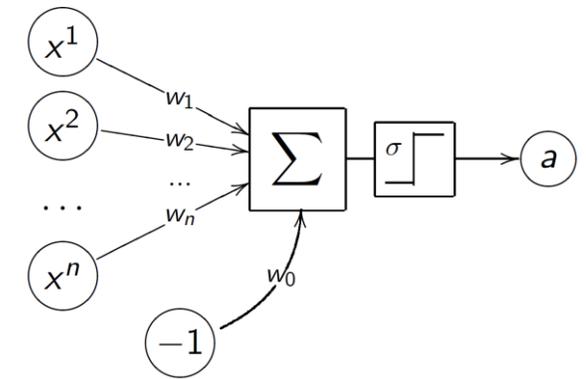
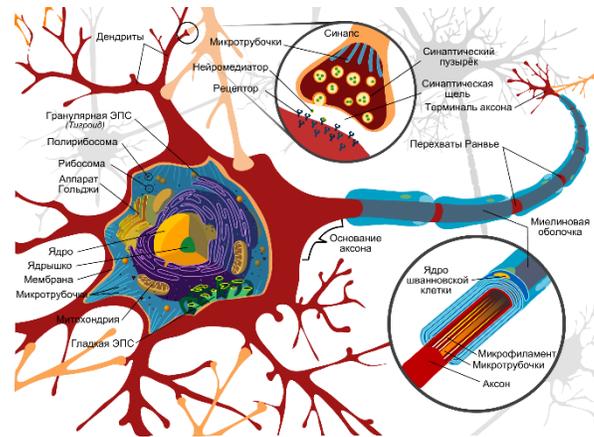
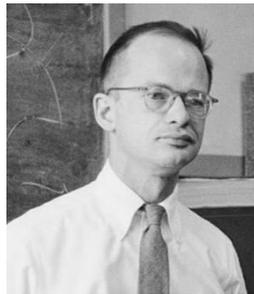
- Предварительная обработка (data preparation)
  - извлечение признаков (feature extraction)
  - отбор признаков (feature selection)
  - восстановление пропусков (missing values)
  - обнаружение выбросов (outlier detection)
  - уменьшение шума (noise reduction)
- Обучение представлений (representation learning)
  - обучение признаков (feature learning)
  - обучение многообразий (manifold learning)
  - анализ главных компонент (principal component analysis)
  - матричные и тензорные разложения (matrix and tensor factorization)

# Типология задач машинного обучения

- Обучение глубоких сетей (deep learning)
- Обучение выявлению связей (relational learning)
- Динамическое обучение (online/incremental learning)
- Обучение с подкреплением (reinforcement learning)
- Активное обучение (active learning)
- Привилегированное обучение (learning with privileged information)
- Обучение с переносом опыта (transfer learning)
- Мета-обучение (meta-learning)

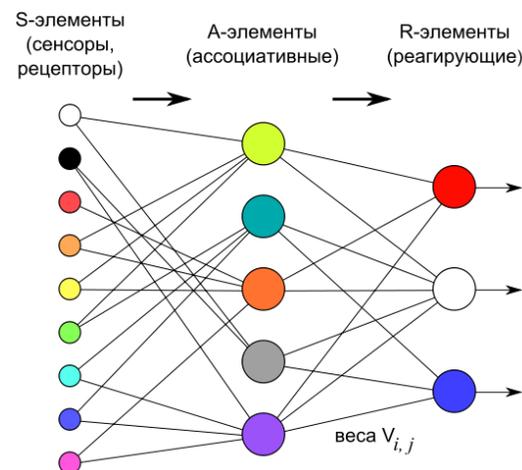
# Что такое «искусственные нейронные сети»

Математическая модель нейрона  
(МакКаллок и Питтс, 1943)

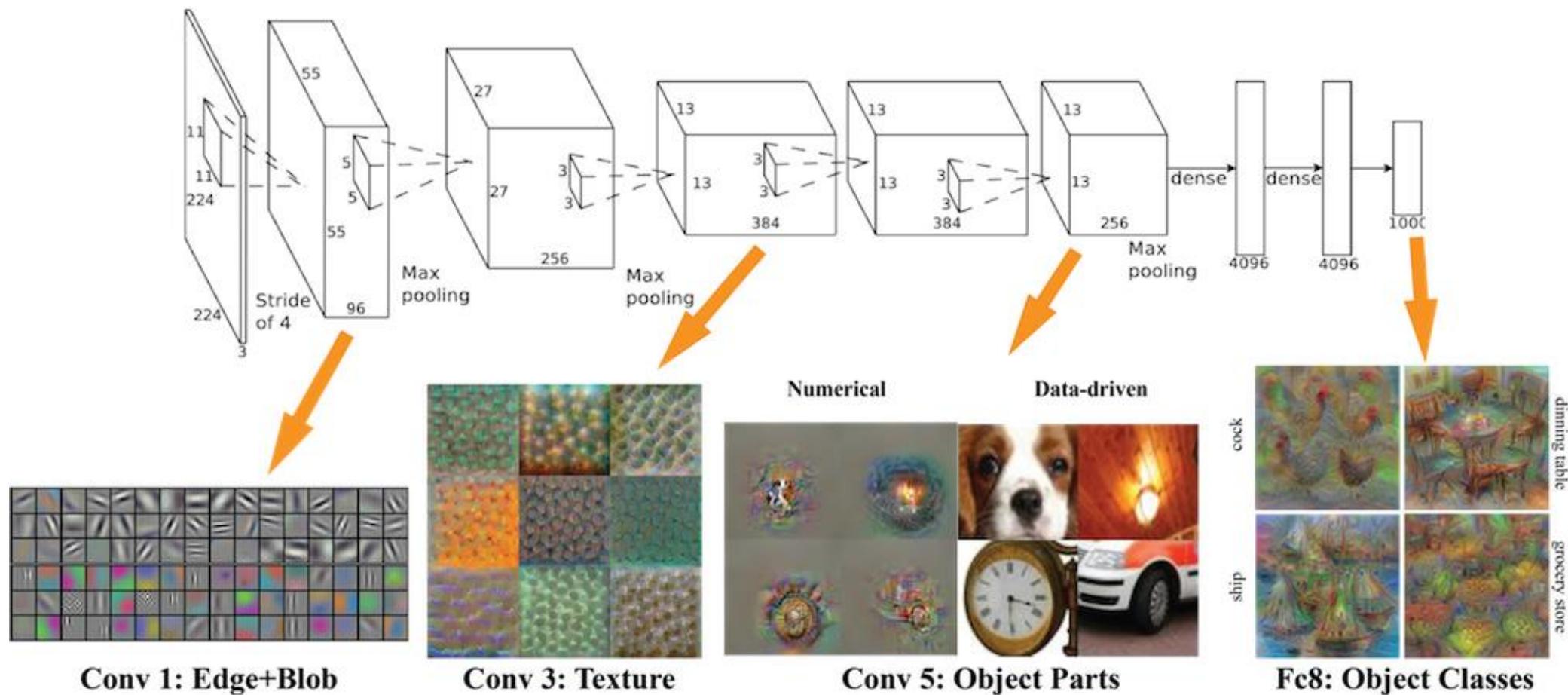


$$a(x, w) = \sigma \left( \sum_{j=1}^n w_j x^j - w_0 \right)$$

Первый нейрокомпьютер Mark-1  
(Фрэнк Розенблатт, 1960)



# Что такое «глубокие нейронные сети»



# Отличия машинного обучения от классического моделирования

- Обилие данных компенсирует недостаток предметных знаний
  - Вместо узких моделей – универсальные аппроксиматоры
  - Нет строгих теорий, но нет и сомнительных модельных гипотез
  - Не все модели машинного обучения – «чёрные ящики»
- Нет чётких различий! Граница размыта!**

*Машинное обучение  
— тоже разновидность  
математического  
моделирования*

# Основные школы машинного обучения

- *Символизм* – поиск логических закономерностей
- *Коннекционизм* – обучаемые нейронные сети
- *Эволюционизм* – адаптивная оптимизация сложных моделей
- *Байесионизм* – оценивание распределений над параметрами
- *Аналогизм* – «близким объектам близкие ответы»
- + *Композиционизм* – кооперация моделей

Педро Домингос. «Верховный алгоритм». 2016.



# Сухой остаток

## Что должен уметь менеджер в области Data Science

- Видеть возможности применения машинного обучения
- Ставить задачи в виде ДНК (Дано-Найти-Критерий)
- Разбираться в методах на уровне «возможности–ограничения»
- Организовывать бизнес-процессы для сбора чистых данных
- Организовывать открытые конкурсы анализа данных
- Запускать пилотные проекты
- Знать экспертное сообщество
- Адекватно оценивать сложность задачи и трудозатраты

# Полезные ссылки

- [ods.ai](https://ods.ai) – *Open Data Science*: русскоязычное сообщество индустрии данных
- [www.MachineLearning.ru](https://www.MachineLearning.ru) – русскоязычная вики
- [www.kdNuggets.com](https://www.kdNuggets.com) – главный сайт датамайнеров
- [www.DataScienceCentral.com](https://www.DataScienceCentral.com) – 72 000 датамайнеров
- [www.kaggle.com](https://www.kaggle.com) – конкурсы анализа данных
- [archive.ics.uci.edu/ml](https://archive.ics.uci.edu/ml) – UCI ML Repository (349 datasets)
- [ru.coursera.org/learn/machine-learning](https://ru.coursera.org/learn/machine-learning) – курс Эндрю Бина
- [ru.coursera.org/learn/vvedenie-mashinnoe-obuchenie](https://ru.coursera.org/learn/vvedenie-mashinnoe-obuchenie) – курс машинного обучения от ВШЭ и ШАД Яндекс
- [ru.coursera.org/specializations/machine-learning-data-analysis](https://ru.coursera.org/specializations/machine-learning-data-analysis) – специализация от МФТИ и ШАД Яндекс

# Рекомендуемая литература

- *Домингос П.* Верховный алгоритм. 2016. 336 с.
- *Коэльо Л. П., Ричарт В.* Построение систем машинного обучения на языке Python. 2016. 302 с.
- *Мерков А. Б.* Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
- *Мерков А. Б.* Распознавание образов. Построение и обучение вероятностных моделей. 2014. 238 с.
- *Воронцов К. В.* Лекции по машинному обучению. [www.MachineLearning.ru](http://www.MachineLearning.ru). 2004-2016.
- *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning. Springer, 2014. 739 p.
- *Bishop C. M.* Pattern Recognition and Machine Learning. - Springer, 2006. 738 p.