



Санкт-Петербургский  
Государственный  
Политехнический  
Университет

Институт прикладной  
математики и механики

КАФЕДРА

ТЕЛЕМАТИКА

Высшая школа ИИ

## курс: Решение прикладных задач методами машинного обучения

### Лекция 4

Технологии решения прикладных задач МО.  
Проблемы «вычисления» решений автоматом и живым  
организмом

---

28 сентября 2022 г.

# Что было на прошлой лекции:

Была рассмотрена тема:

- **Проблемы «решения» прикладных задач методами МО. Было показано, что решения прикладных задач методами МО сводятся к**
  - **сбору конечного множество прецедентов (объектов, ситуаций), каждому из которых сопоставляются данные «правильные решения»**
  - **совокупность всех частных описаний прецедентов называется обучающей выборкой.**
  - **задача МО сводится к тому, чтобы по данным обучающей выборки выявить общие зависимости, закономерности, взаимосвязи, присущие не только этой конкретной выборке, но вообще всем прецедентам, в том числе тем, которые ещё не наблюдались.**

Вопросы, на которые надо дать ответы:

1. **возможно ли решить такую задачу в принципе ? и если «да» , то....**
  1. **можно ли «доверять» полученному решению ?**
  2. **как доказать, что решение верное ?**
  3. **в какой форме можно передать вы ?**

- **Тема этой лекции: Технологии решения прикладных задач МО**

- **Перспективы: технологические, футурологические...**

# ЭВОЛЮЦИЯ ТЕХНОЛОГИЙ ОБУЧЕНИЯ

- Обучение – особый «инструмент» решения задач. Этот инструмент выявления «закономерностей» между входными данными и решением. Этот инструмент используется, когда: 1) нет алгоритма вычисления решения по готовой формуле, 2) нет статистических или причинных связей, между входными данными и решением.
- В процессе обучения человека ему **передаются «знания»**, позволяющие не только «вычислить» решение, но и проверить его «правильность».
- В процессе же «обучения» машин на их вход машины **передаются данные**, которые обрабатываются с помощью вычислений таким образом, чтобы правильность полученного результата мог проверить человек.
- С точки зрения теории информации, проводимые машиной **вычисления — это процесс формирования сообщения** (носителя новой апостериорной информации) из имеющихся данных . Для того, чтобы сделать компьютер «умнее» надо ответить на вопросы:
  - **кому** предназначено «вычисленное» сообщение ?
  - **как и кто это** сообщение может прочесть и в дальнейшем использовать?

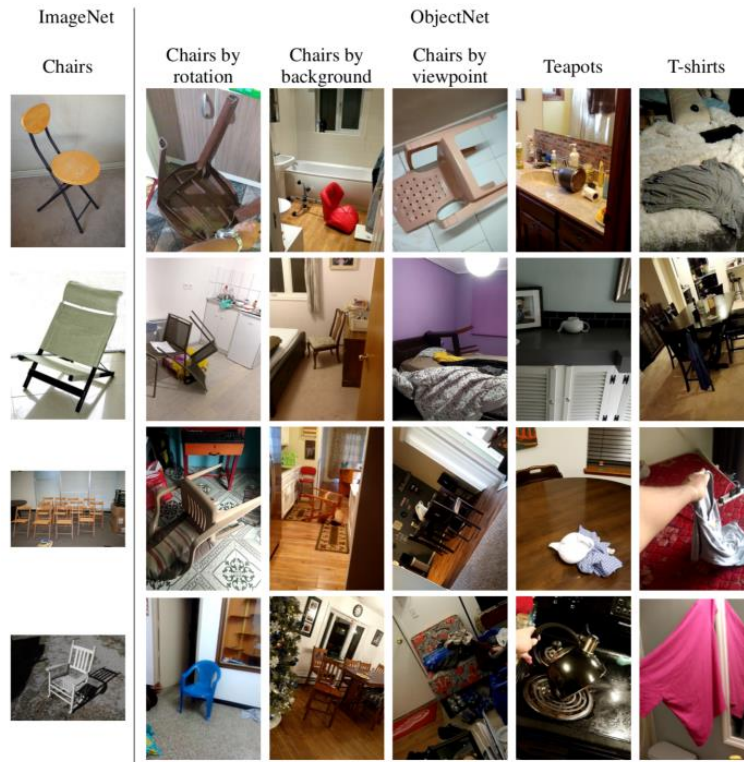
**а также**

  - **надо ли** ускорить процесс вычисления сообщения ?
  - **может ли** вычисленное сообщение содержать ошибки ?
- **Ключевой вопрос технологий машинного обучения:** как интерпретировать результаты вычислений, используя для этого технологии и методы.... апостериорных вероятностей и статистических выводов и т.н. **glorified statistics ....**

# Технологии передачи: знаний vs информации vs сообщений

- Технология передачи знаний основана на том, что знания нельзя передать, а знаниям можно получить. Для этого «приемник» должен обладать первоначальным набором слов, которые используются для записи знаний. Слова для записи знаний образуют конечное множество и хранятся в специальном словаре – тезаурусе. В этом словаре все слова используемого го языка представлены максимально полно и с исчерпывающим перечнем их употребления в тексте. Передача знаний происходит, когда меняется объем тезауруса, число данных в процессе передачи знаний м.б. «небольшой» по объему. Изменение тезауруса приводит к изменению знаний «приемника»
- Технология передачи сообщений основана на том, что объем передаваемой информации целиком содержится в этом сообщении. Объем информации оценивается через логарифмическую функцию от вероятности получения этого сообщения.

# «Проклятие» отождествления: словарь образов и тезаурус понятий



Объекты в обучающих наборах данных по сравнению с объектами в реальном мире . По мере того, как среда становится все более сложной, практически невозможно охватить все варианты распределения посредством добавления большего количества обучающих примеров. Взаимодействовать с миром машинам проще через «образы», чем с помощью понятий.

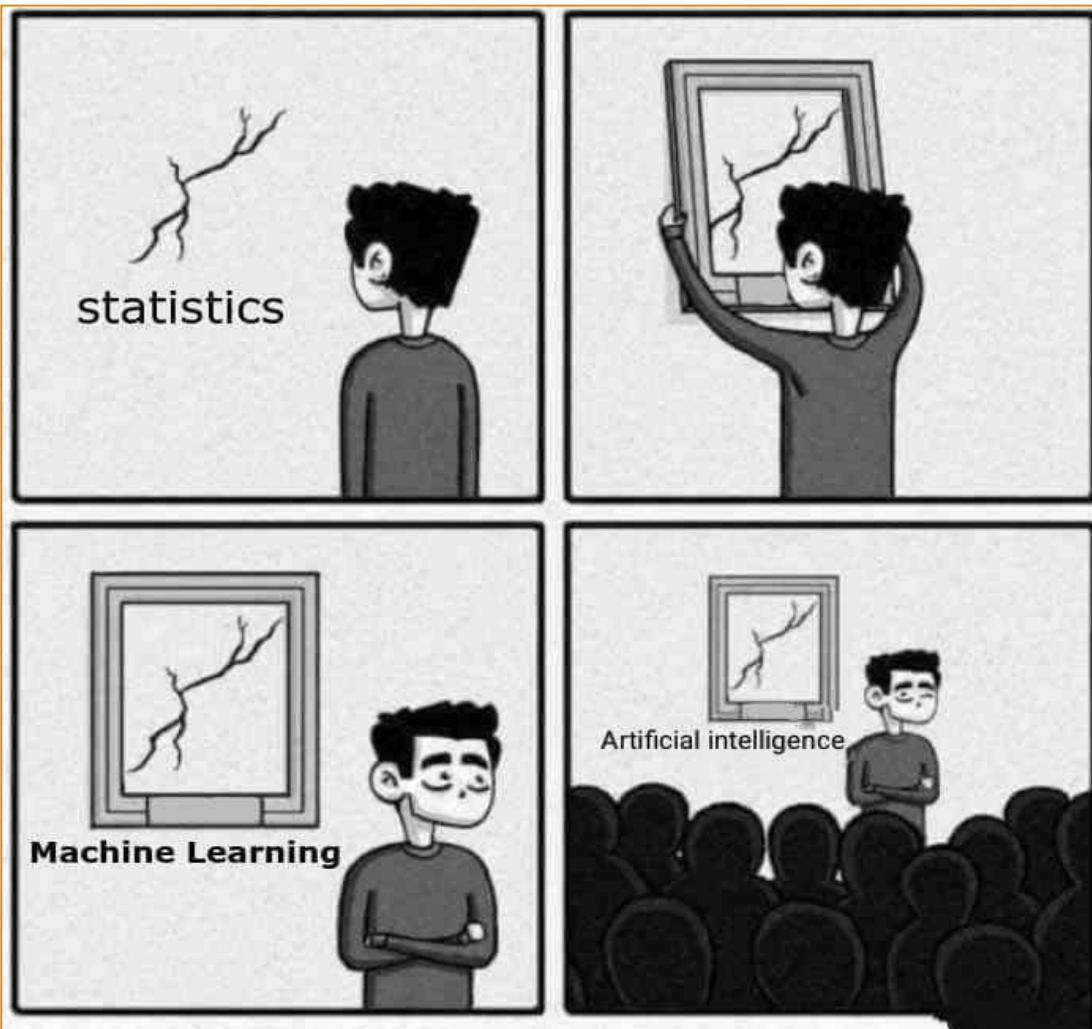
# Пропедевтика (с чего надо начать) машинного обучения

**Что надо знать о методах обучения.** Истинность любых дедуктивных (формальных) методов обучения основана на непротиворечивости используемой аксиоматики. Поэтому обучение надо начинать с формирования «базиса» или задания конечного множества классов объектов, рассматриваемых как «алфавит понятий» предметной области.

- Для достижения практически значимого результата с использованием индуктивных методов или с помощью **алгоритмов машинного обучения их** разработчики должны «задать» описание предметной области с помощью конечного базиса классов. При этом согласно теоремам Геделя применение «формальных» методов, делает процесс обучения противоречивым и, главное, «неполным».
- Поэтому, стандартные модели обучения следует «пополнить» с учетом :
  - Причинно-следственных связей
  - Топологических инвариантов
  - Субъективного тезауруса того, **кто воспринимает данные**

# glorified statistics ...МЕТОДЫ ПРЯМОГО ВЗАИМОДЕЙСТВИЯ С ДАННЫМИ,

счетное множество слов vs континууму понятий



**Сила обучения не в слова, а в их визуализации.**

- Число слов, которые можно напечатать на компьютере образуют счетное множество.
- Число понятий, которые можно представить в сознании, образуют множество континуума.

«Большие данные», которые нужны для машинного обучения – это такие данные, объем которых не помещается в «персональный компьютер» вашего мозга.

## Уточнения (1)

Вопрос «В чем разница между машинным обучением и статистикой?» задается уже несколько десятилетий.

Ответ может такой: **Машинное обучение — это статистика, масштабируемая до «больших данных»**. Итак

- и статистика, и машинное обучение создают модели на основе данных, но для разных целей.

Но статистики в значительной степени

- сосредоточены на использовании особого типа метрик (**описательных** — «среднее», «дисперсия» и **выводных** — ... сколько людей нуждаются в медицинской помощи), которые обеспечивают форму сокращения данных ...для их лучшего «понимания».
- обеспечивают основу для **принятия эффективных (минимум дисперсии) решений**, позволяя оценить вероятности, связанные с различными исходами.
- занимаются достоверностью модели, точной оценкой параметров модели и выводами из модели.

## Уточнения (2)

- В машинном обучении основной задачей является **прогностическое моделирование** (прогнозная аналитика): создание моделей с целью **прогнозирования дескрипторов** новых наборов данных.
  - Суть МО в том, что :
    - 1) имеется набор ранее помеченных примеров.
    - 2) каждый пример имеет метку, которая, в зависимости от типа задачи, может быть либо
      - именем класса (классификация), либо
      - числовым значением (регрессия).
    - 3) строится Алгоритм обучения , который
      - анализирует наборы размеченных данных и **создает оператор**, который для нового набора примеров может довольно точно предсказать его имя класса или числовое значение
      - некоторая часть - контрольный набор используется для проверки правильности работы **созданной процедуры**
- При этом предполагается, что будущее будет похоже на прошлое. Т.е. причинно-следственные связи подтверждены и сохраняются.

## Уточнения (3)

- Девиз машинного обучения может быть таким: **доказательство адекватности модели находится в тестовом наборе.** В результате:
    - производительность тестового набора является арбитром качества модели, которая м.б. как параметрической, так и не параметрической
    - в случаях, когда предположения нарушаются, существует ненулевая вероятность, когда разработанная модель может успешно работать,
    - Методы МО обладают робастностью, поэтому продолжают работать, даже когда нарушаются предположения относительно класса распределений (гауссовское, пуассоновское, биномиальное...).
    - Однако, предполагается, что выборки обрабатываемых данных выбраны независимыми и одинаково распределенными, поэтому обучающие выборки являются репрезентативными для всей популяции.
    - Решение должно вычисляться быстро, но... не обязательно точно
- См. <https://www.kdnuggets.com/2017/06/regression-analysis-really-machine-learning.html>
- (подготовить доклад: что отличает статистику от машинного обучения)

# Можно ли объяснить «машине», что такое турбореактивный двигатель ?

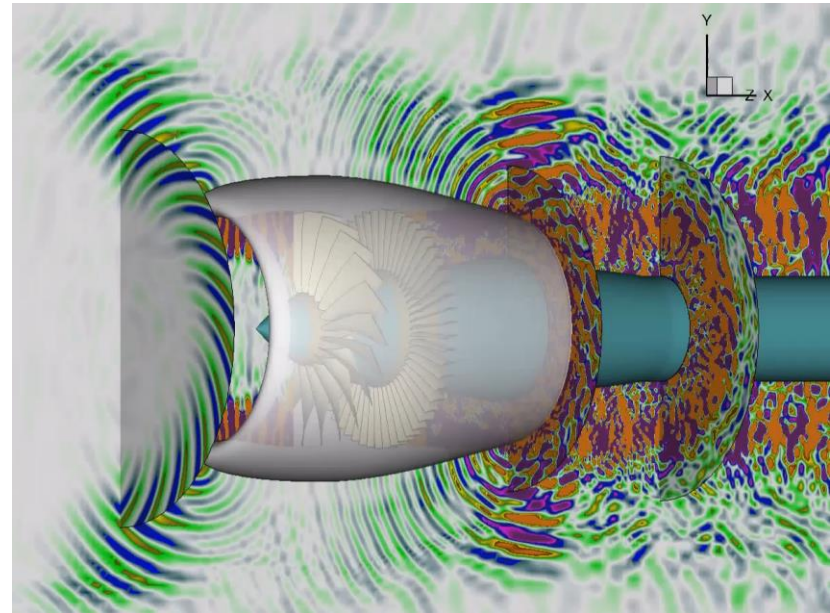


Задача: прямой расчет и визуализация акустических характеристик турбореактивного двигателя

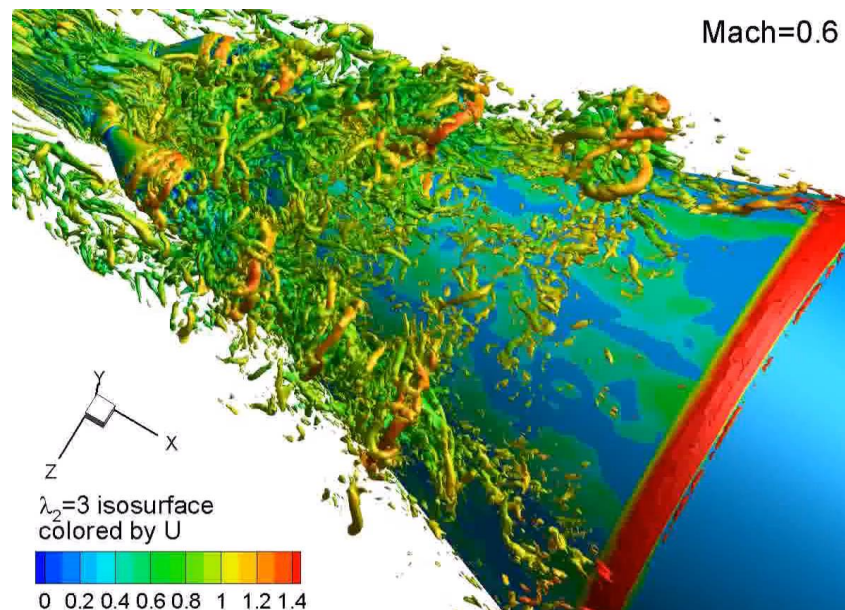
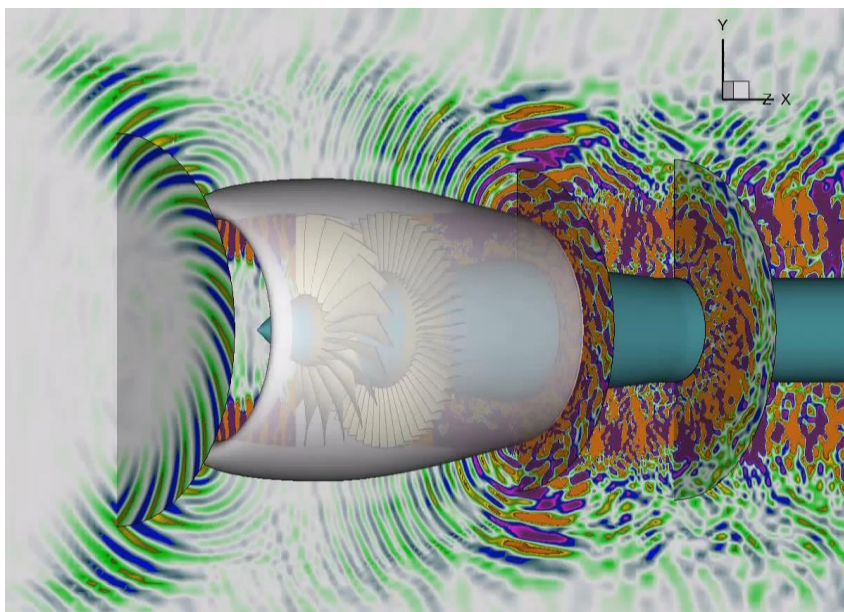
- **Возможно ли использование** в качестве обучающей выборки динамического набора данных, полученных в результате вычислений на основе уравнений математической физики ?
- **Извлечение** объясняющих результатов из набора данных моделирования, полученных в процессе прямых вычислений. Пример, обучение на основе данных **визуализации протекающих процессов**

Физическая модель турбореактивного авиационного двигателя

Анимация — пример **суррогатной модели** интерпретации результатов расчета звуковых волн, излучаемых турбореактивным двигателем



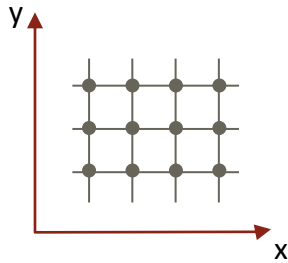
«Понять нельзя вычислить»: как правильно в этой фразе расставить запятые ?



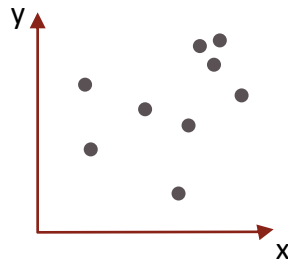
Анимация вихревых структур и пульсаций давления на поверхности возвращаемого космического аппарата в процессе выведения

# Решение задач обучения на основе данных «расчетов» с использованием «первых принципов» и классов эквивалентности

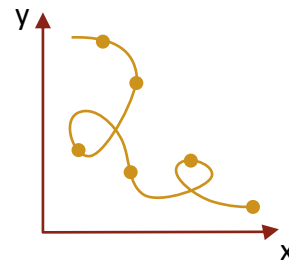
типовые расчеты на основе «первых принципов ...»



регулярная сетка



случайный поиск



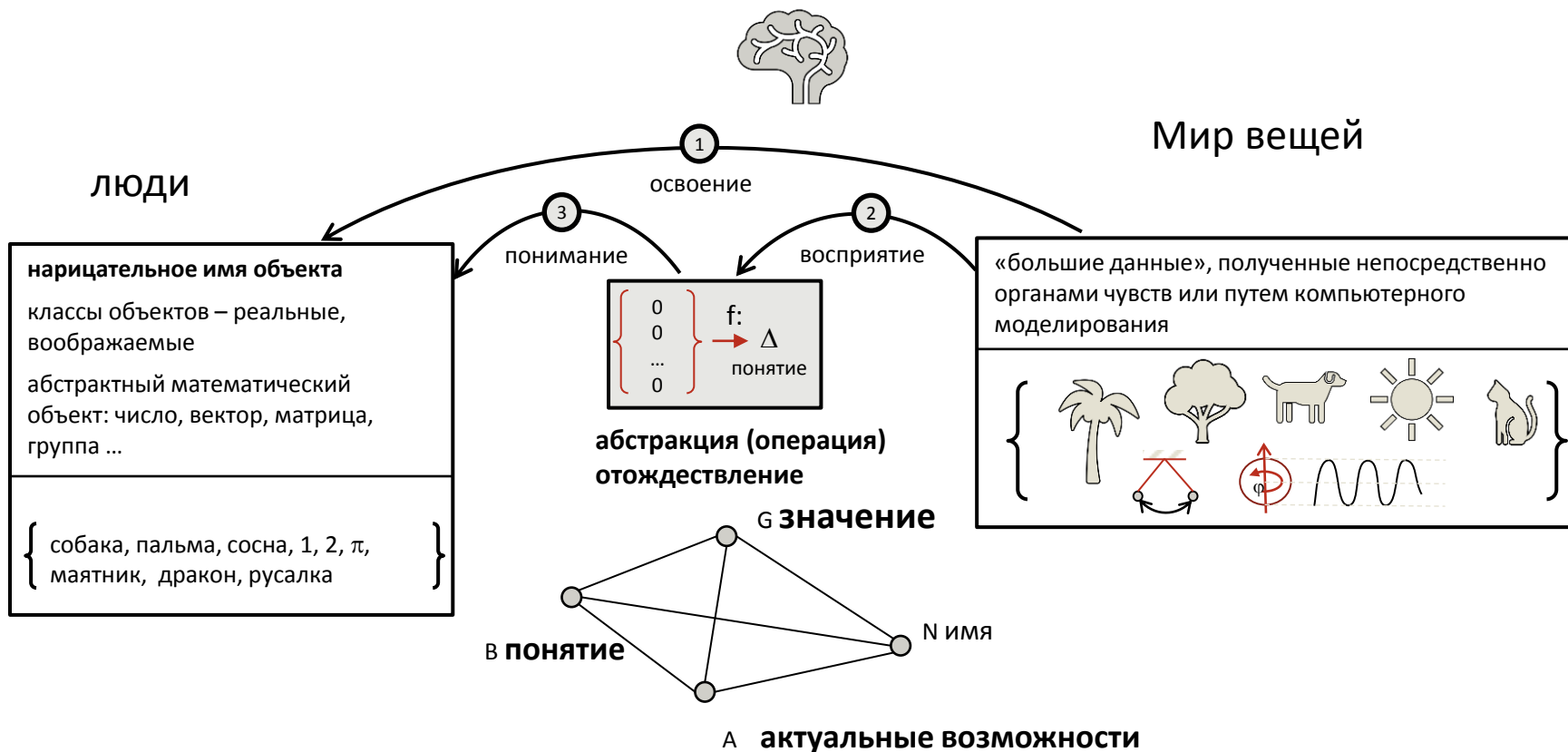
градиентный «спуск»

накопление «опыта» путем формирования классов эквивалентности (аксиома отождествления)

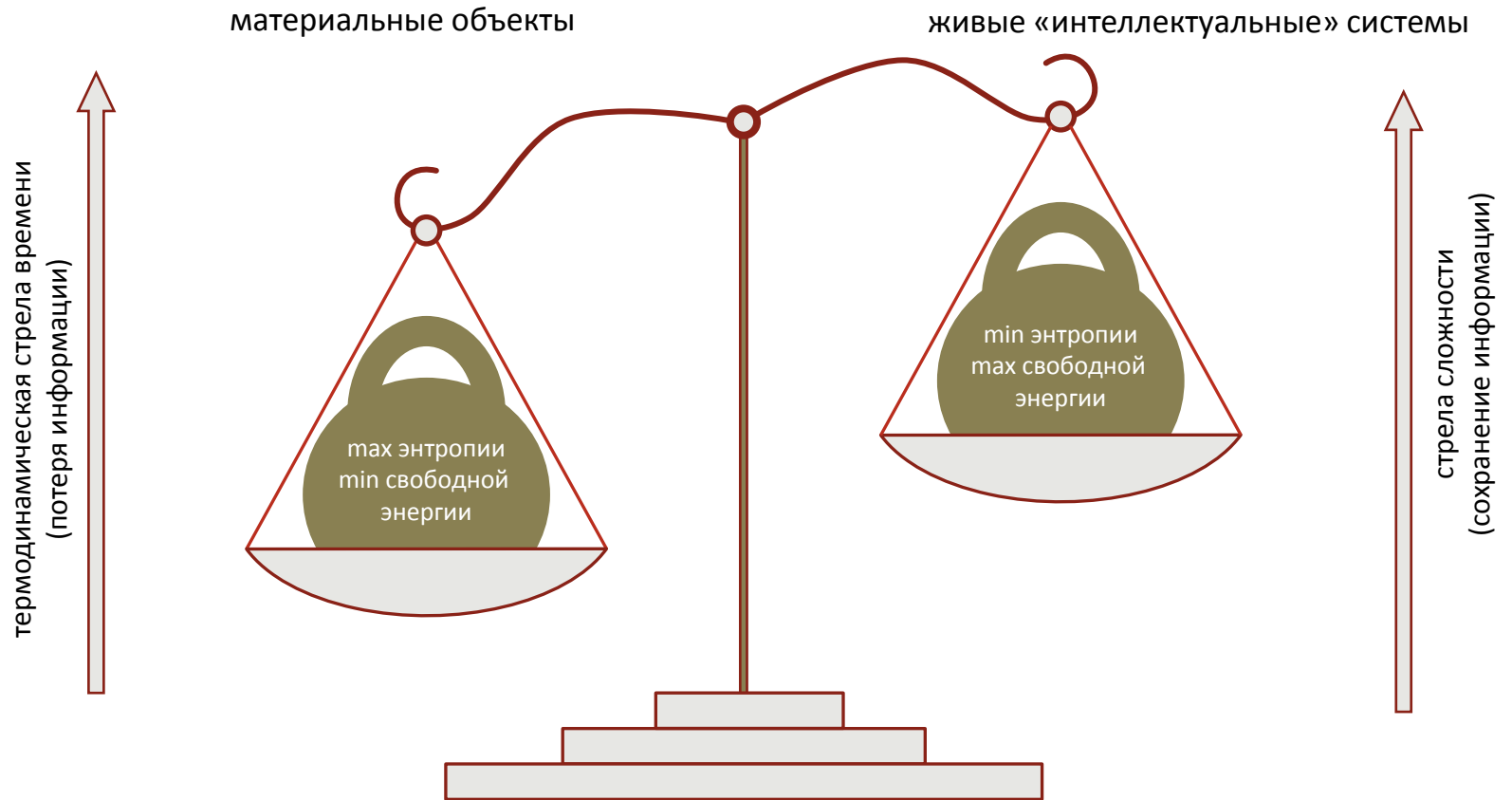
$$P(A|B) = P(B|A)P(A)/P(B)$$

# Аналогии из мира живых организмов vs машин vs людей

Мозг животных не использует функцию «понимания» через абстрагирование, а реализует мировосприятие, мироосвоение через



# Устойчивое неравновесие «восприятие- понимание»



# Ключевые слова, входящие в тезаурус

**Причина – основание** для каких-нибудь действий

**Интенция** – (лат. intentio «намерение, стремление») — **направленность** сознания, мышления на какой-либо предмет

**Когнитивные функции** - термин используется для описания умственных процессов в головном мозге, благодаря которым человек получает возможность воспринимать, передавать, анализировать и запоминать различную информацию ,

**Роль окружающей среды** в продуцировании поведения – зависит от того, как обрабатывается в организме поступающую к нему информация.

**Онтогенез** - способы взаимодействия генетической информации и информации, полученной из окружающей среды (или путем научения), которые формируют соответствующий тип поведения «живого организма».



# Проблема «вычисления» состояний в пространстве возможностей

**Суть:** Почти все живые организмы (компьютеры) построены из одних и тех же веществ (компонент), причем различия между видами организмов ( компьютеров) определяются различиями в способах объединения этих основных компонент.

**Проблема:** Понять механизмы, контролирующие объединение отдельных клеток в живой организм. Какая информация в генах контролирует процесс развитие организма, управляя синтезом белков, и организацию процессов онтогенеза в целом. **Вопросы с точки зрения методов обучения**

- Какая информация должна быть передана в «генном» сообщении, чтобы из отдельных белков собралось «здание организма»?
- Какая информация нужна от внешней среды ?
- Что нужно сделать, чтобы генетический код стало частью «процесса онтогенеза» ?

# Начало процесса обучения машин:

## метод Корсакова (1832)

- С.Н. Корсаков. Статья : «Начертание нового способа исследования при помощи машин, **сравнивающих идеи**».
- Цитата: «Человек мыслит, но действия его носят механический характер: он приказывает, и его ноги идут, а руки двигаются.
- Разум занимается тем, что действует через формирование понятий, реальные изменения присущи только материальным объектам

# «Первая перфокарта» Корсакова как код «идеи»

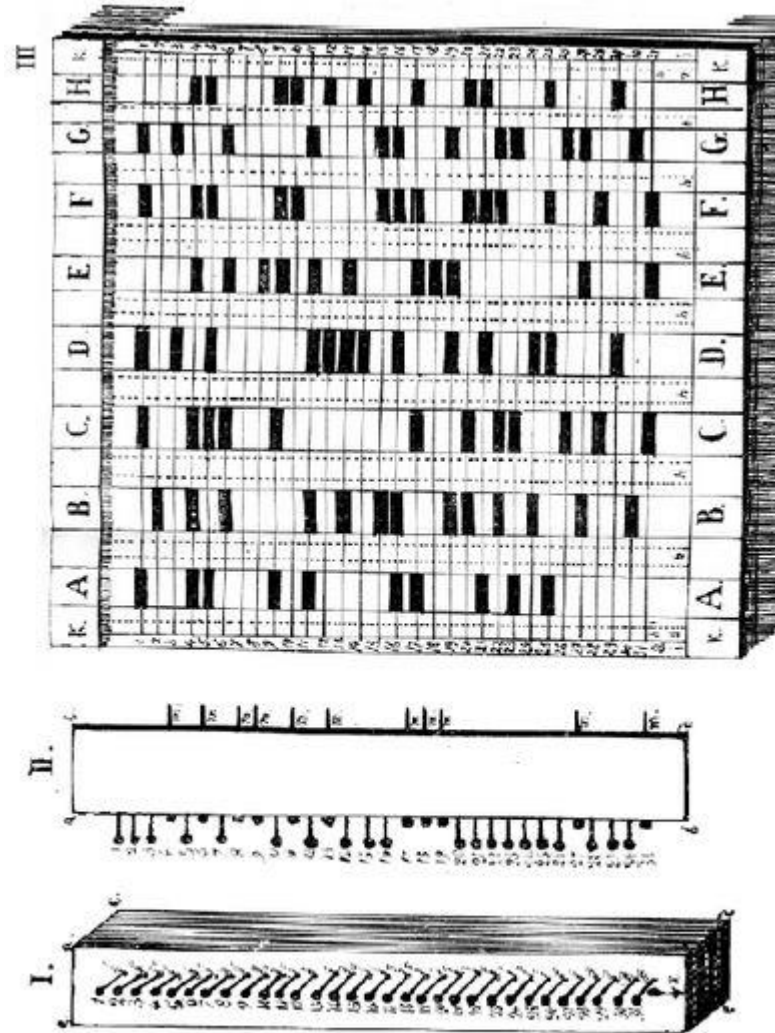


Рис. 1. Прямолинейный гомеоскоп с неподвижными частями (I, II); гомеоскопическая таблица (вид сверху) (III)

Может ли машина,  
объяснить то, что  
она делает ?

Да, если владеет  
языком понятий.

Что необходимо,  
чтобы овладеть  
языком понятий ?

# Выводы

- Чтобы физическая система была бы способна к обучению или восприятию знаний, она должна обладать исходной базой первичных понятий – алфавитом понятий
- Результаты обработки данных о состояниях физической системы могут играть как информационную, так и каузальную (причинную) роль в движении системы
- При определенных условиях можно ли считать, что информация это и результат и причина движения