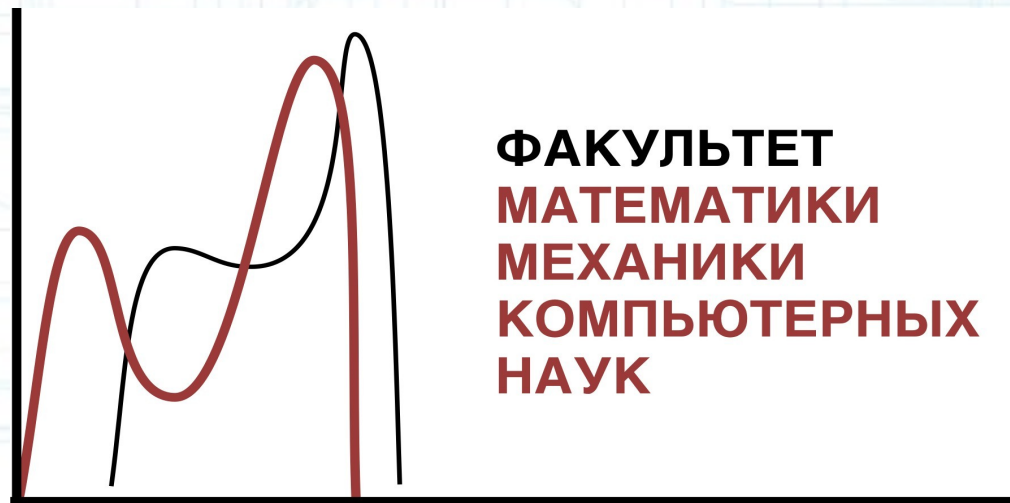


Машинное обучение и интеллектуальный анализ данных

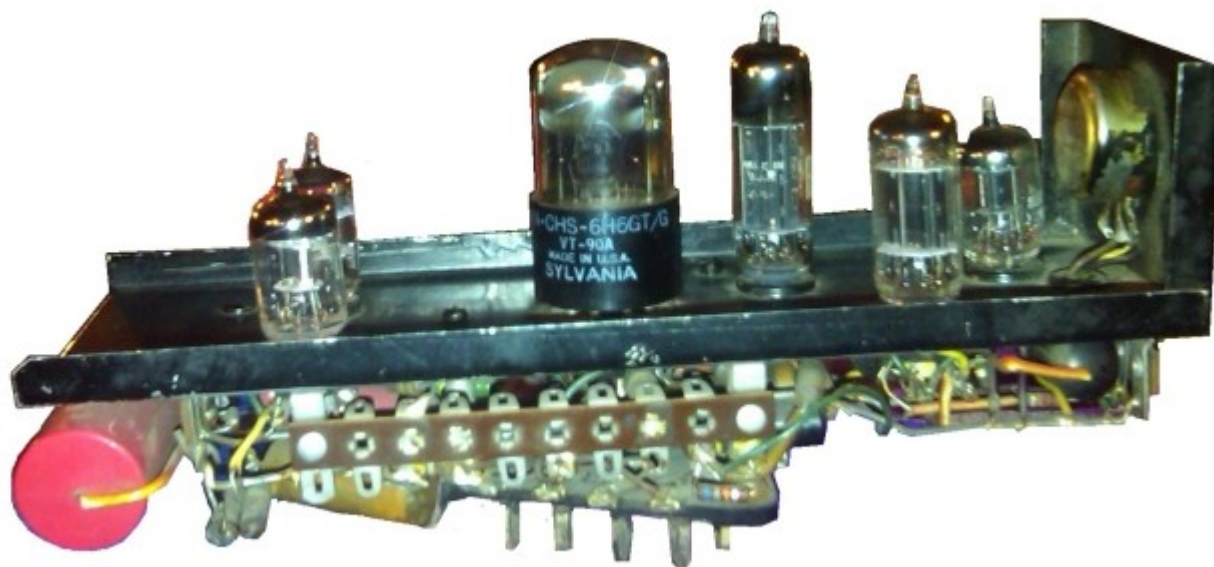
История



Содержание

- История
 - Кибернетика и первые нейронные сети
 - Первые цифровые компьютеры
 - Появление экспертных систем
 - Вероятностный подход
 - Случайный лес
- Последние достижения

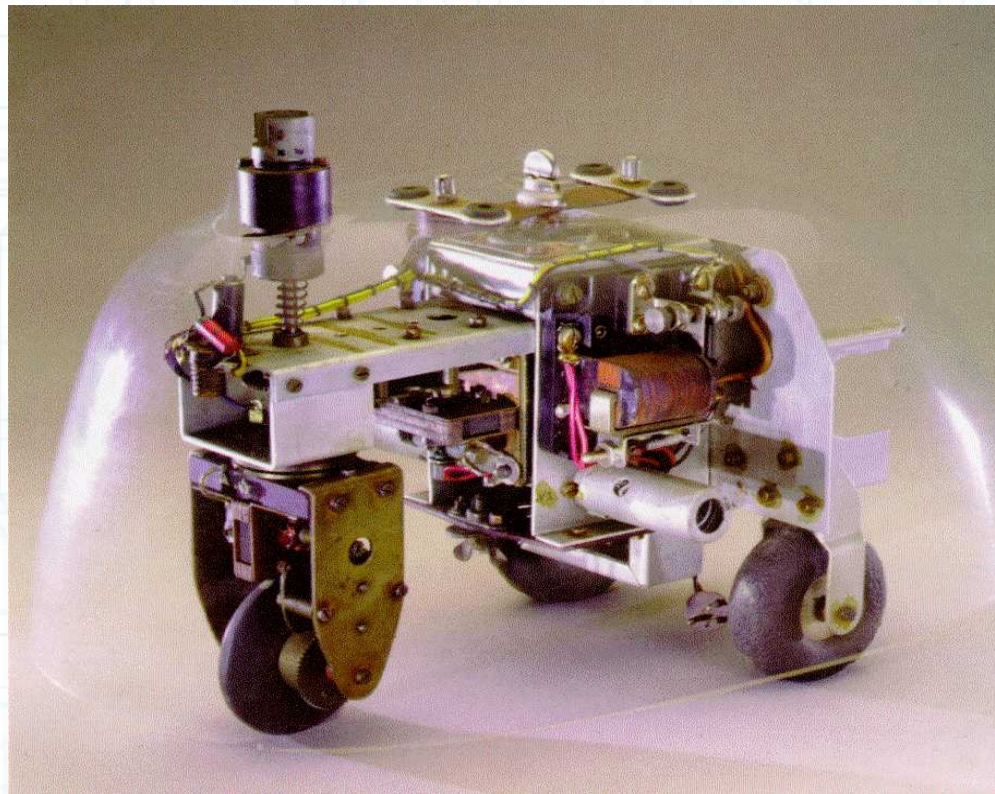
Кибернетика и первые нейронные сети



SNARC - Marvin Lee Minsky

Кибернетика и первые нейронные сети

- Черепаха (конструктор Grey Walter, 1949)



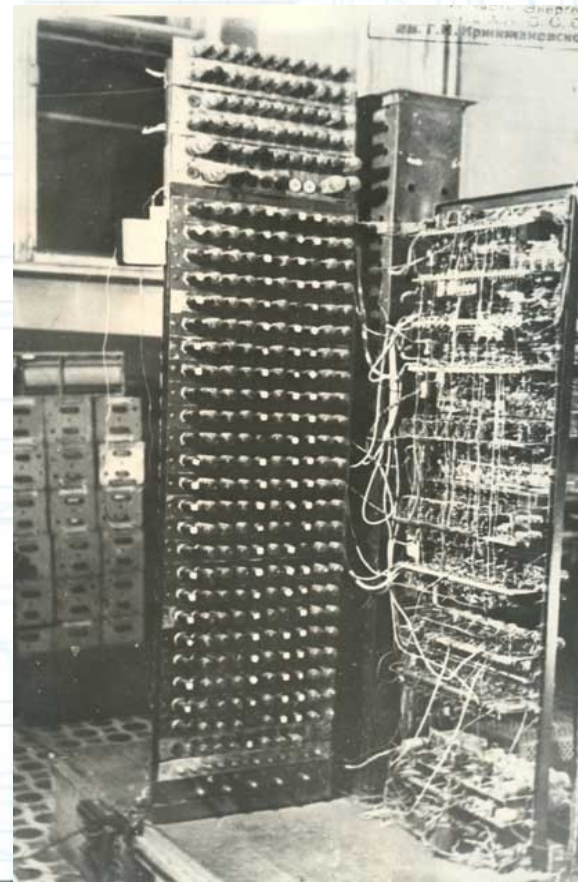
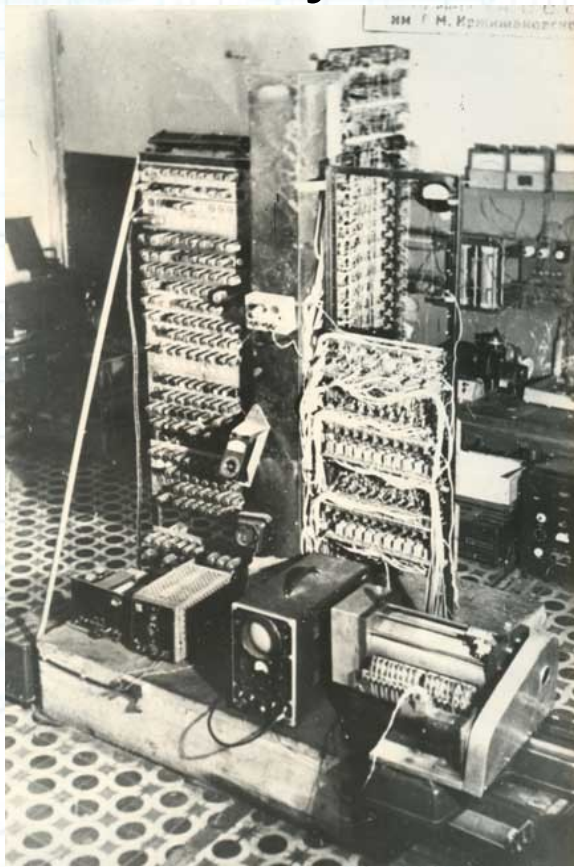
Кибернетика и первые нейронные сети

- Цифровая электронная вычислительная машина Б.И. Рамеева и И.С. Брука
- Решала дифференциальные уравнения до шестого порядка
- Авторское свидетельство выдано 4 декабря 1948

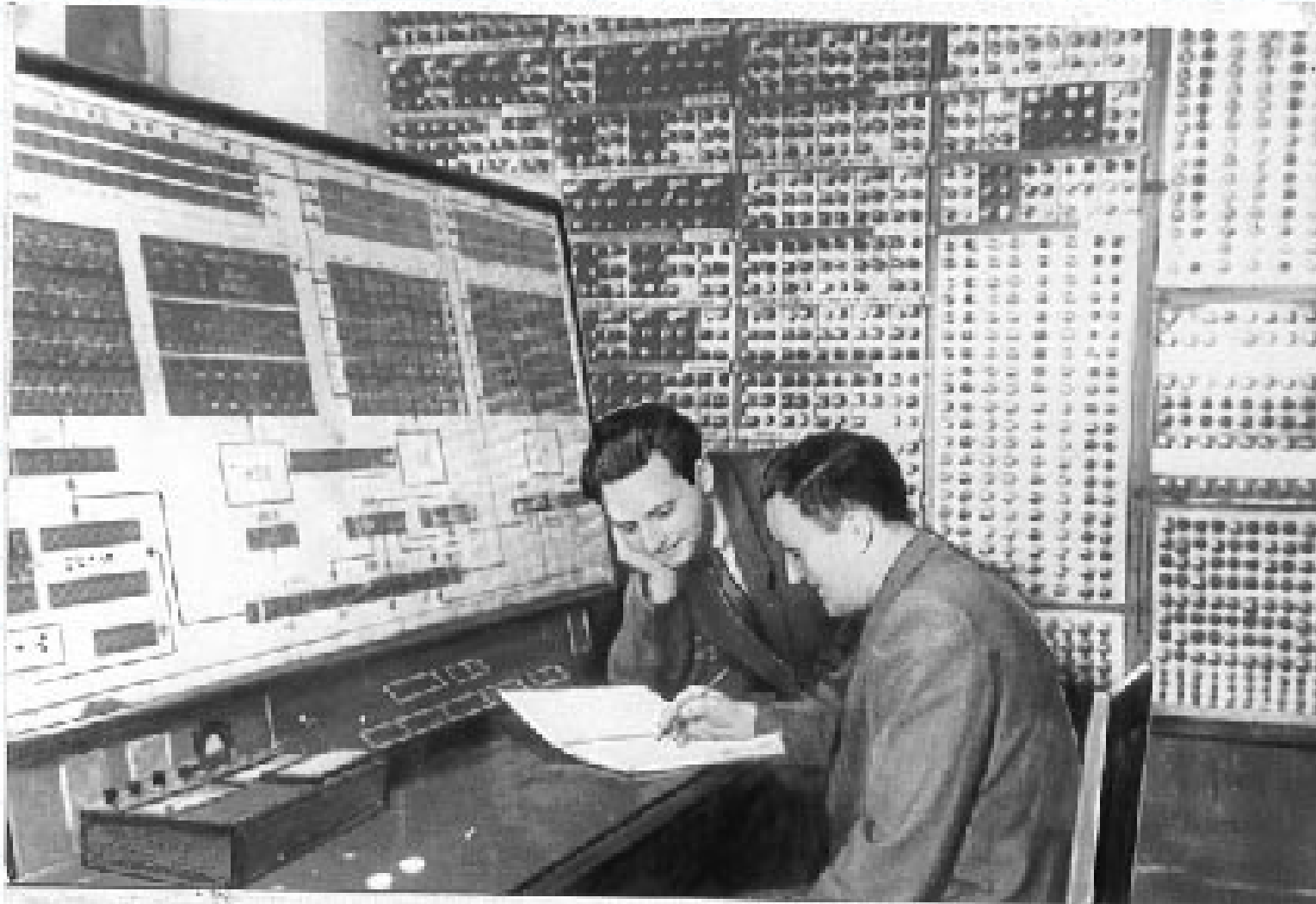


Первые цифровые компьютеры

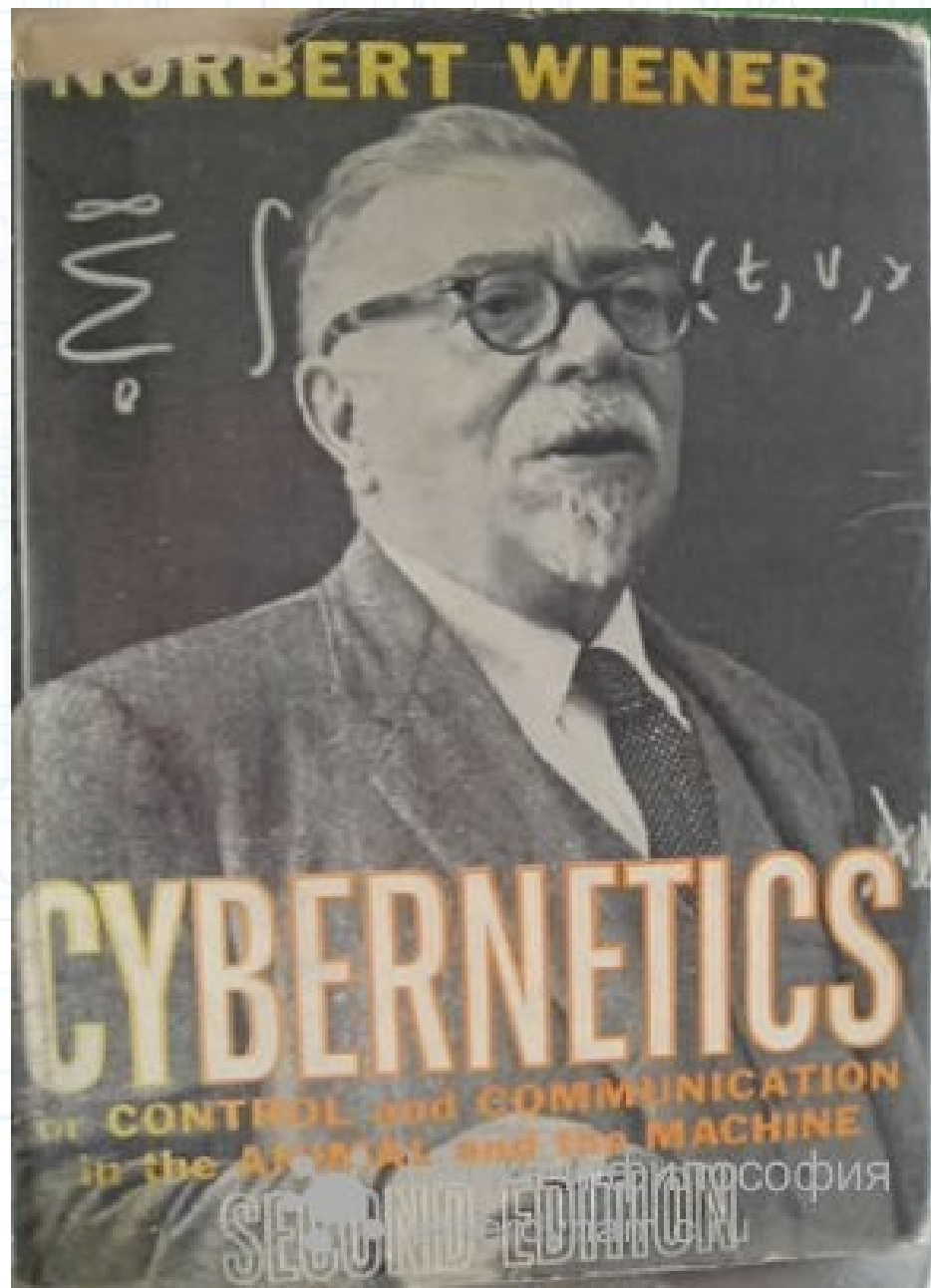
- ЭВМ М1 - первая в мире машина на полупроводниковых приборах
- Построена под руководством И.С.Брука в 1951 году



Первые цифровые компьютеры



1948г. “Кибернетика” Н.Винера

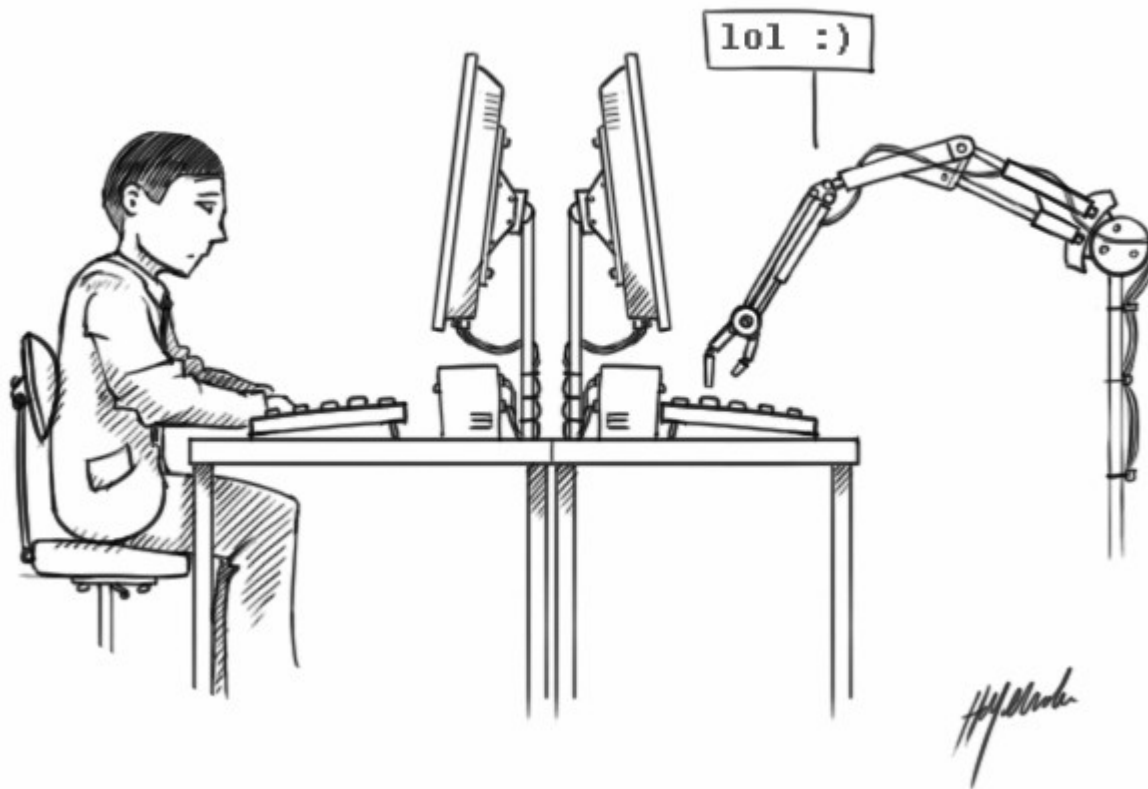


Алан Тьюринг

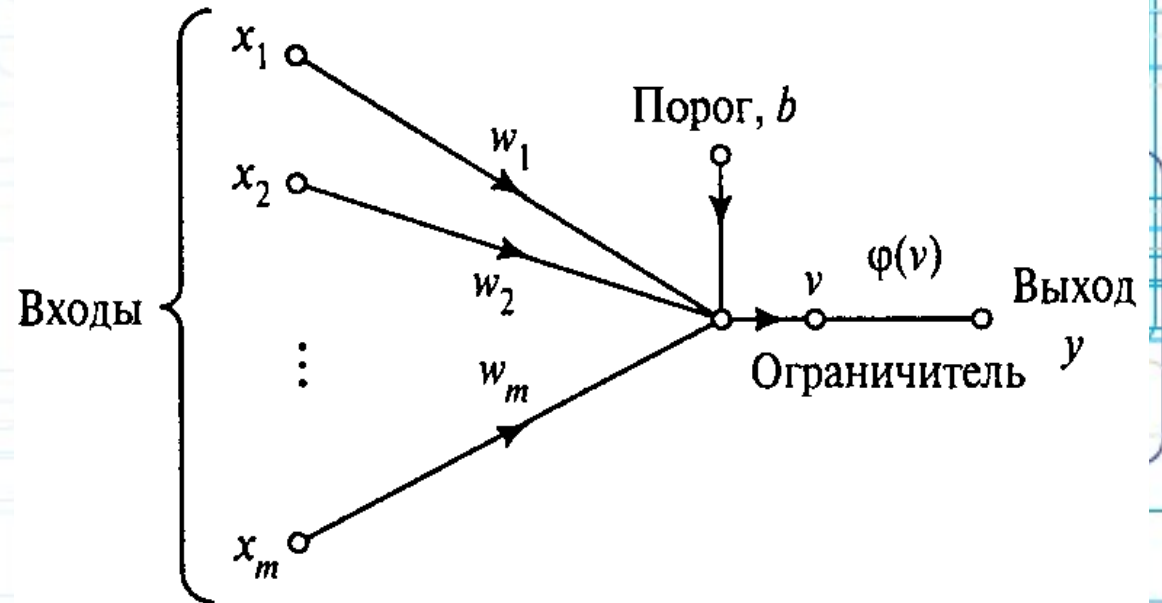
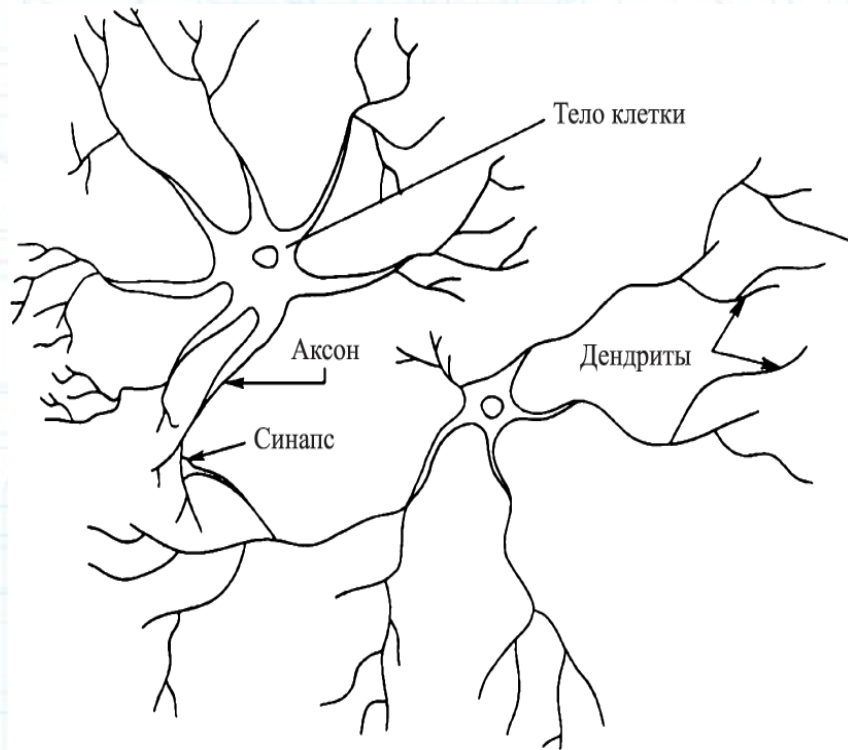
**Может ли
машина
мыслить как
человек?**



Тест Тьюринга



Перцептрон Розенблатта



“Шашки” Артура Самуэля



1956г. Дартмутский семинар

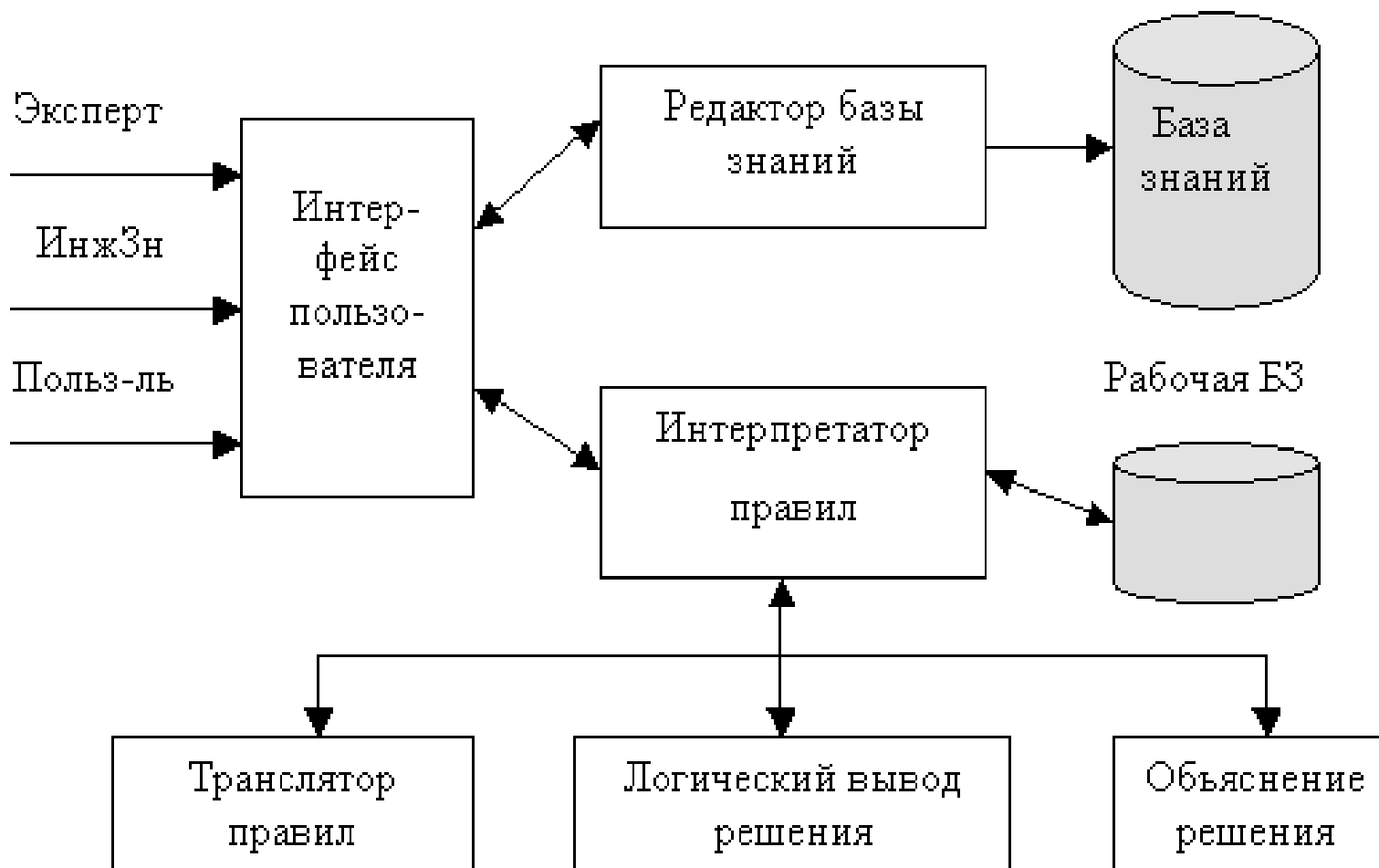
- Целью конференции было рассмотрение вопроса: можно ли моделировать рассуждения, интеллект и творческие процессы с помощью вычислительных машин
- Утвердила появление новой области науки и дала ей название — «Artificial Intelligence»

1974–1980 – зима для Искусственного интеллекта

- **Наивный оптимизм:**
 - 1954 через 3 — 5 лет задача машинного перевода будет полностью решена
 - 1958 через 10 лет компьютер станет чемпионом мира по шахматам; докажет важную математическую теорему
 - 1970 через 3 — 8 лет будет создан искусственный интеллект общего назначения, сравнимый с интеллектом среднестатистического человека
- **парадокс Моравека**

Первые экспертные системы: Dendral и MYCIN

Эдвард Фейгенбаум

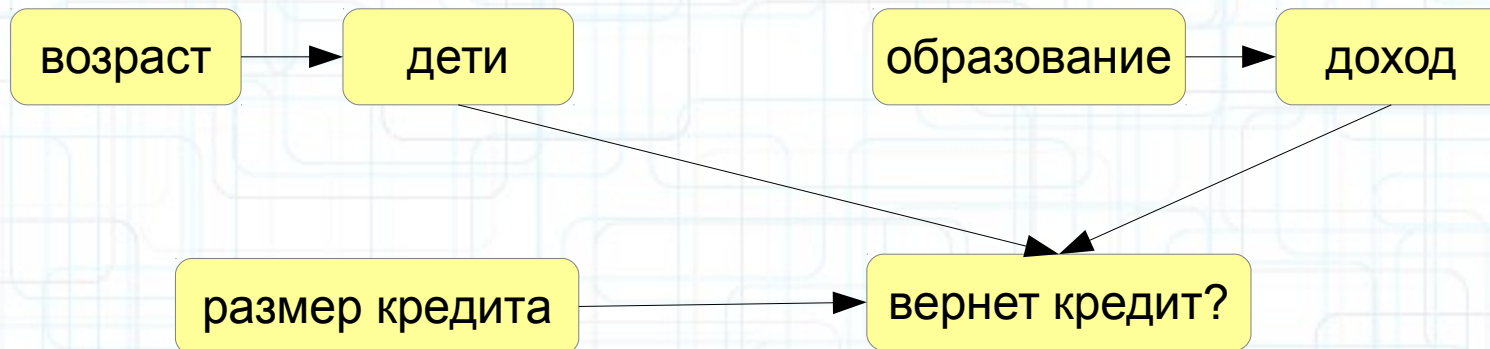


1980–1987 внедрение экспертных систем

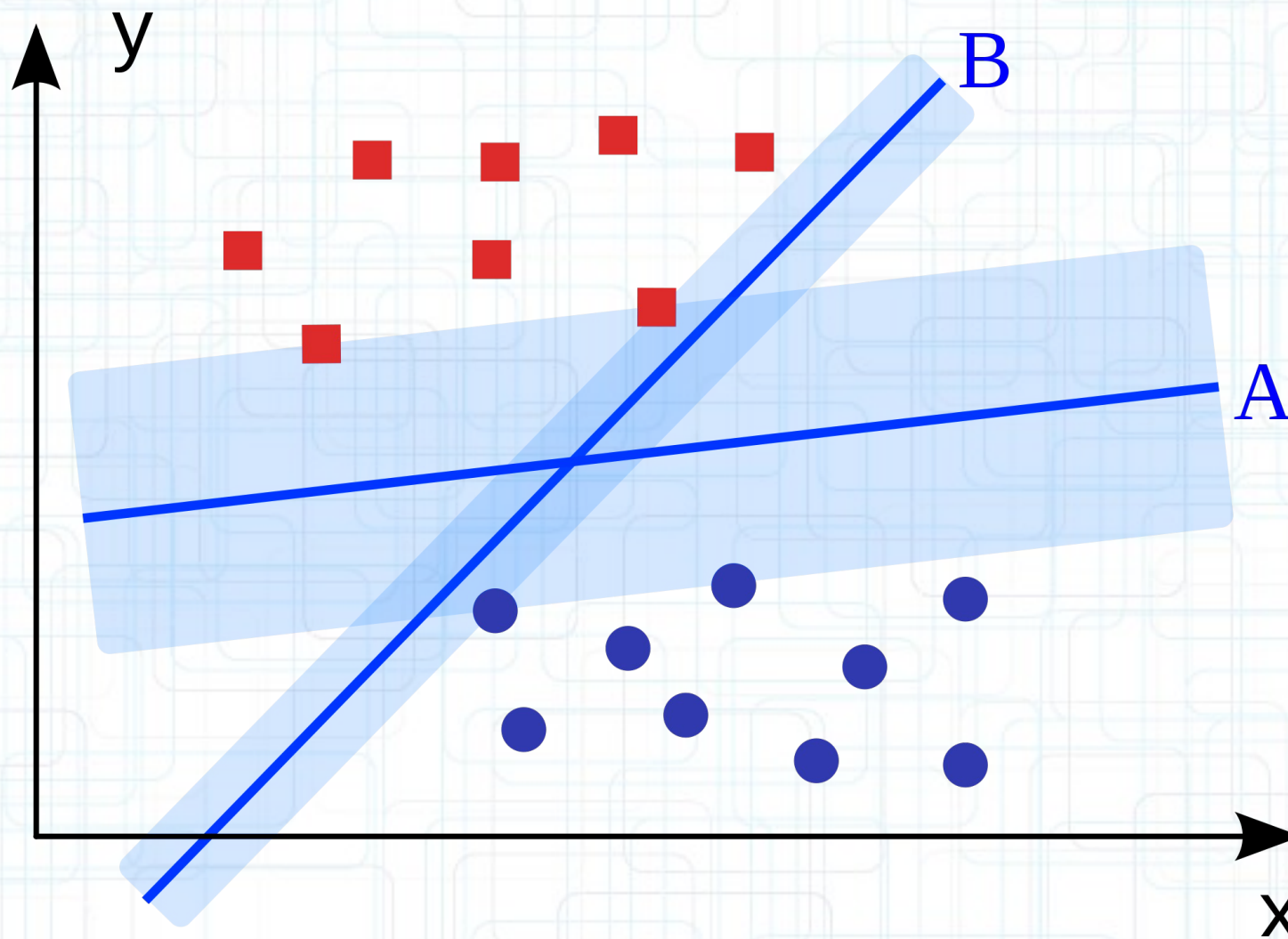
- XCON сэкономила фирме DEC 25 млн \$ в год
- PROSPECTOR для геологической разведки месторождений полезных ископаемых
- Количество экспертных систем в СССР исчисляется сотнями (обзор в справочнике под. ред. Д.А.Поспелова)

Байесовские сети доверия

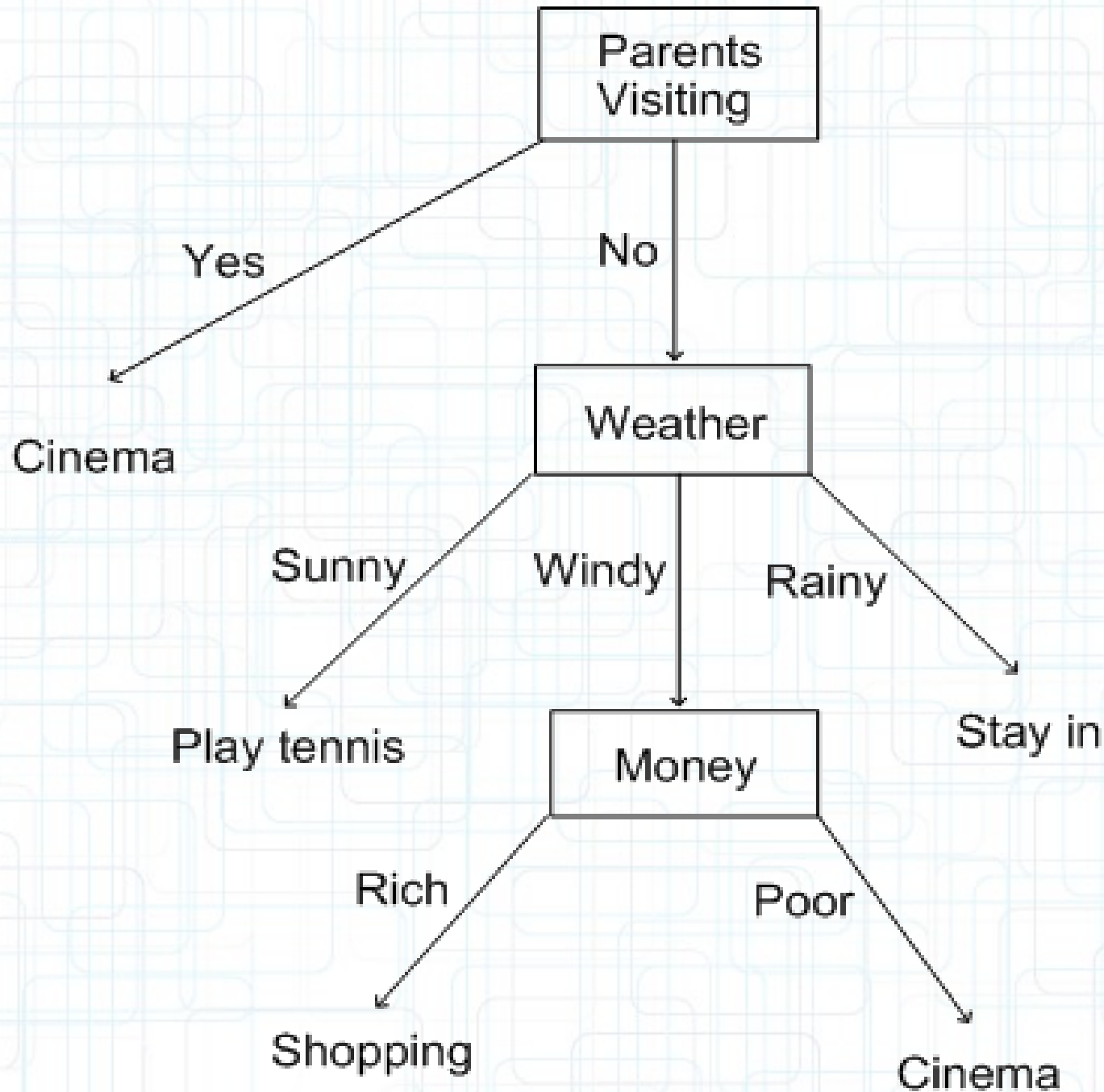
- Решение задач ИИ основывается не на знаниях экспертов, а на статистических данных
- Джуда Перл создает аппарат байесовских сетей доверия для задач классификации (1988 г.)



1995 К.Кортес и В.Вапник изобретают SVM



Деревья решений



Деревья решений

- Поиск наилучших закономерностей
- Группа 100 пациентов с подозрением на менингит (вероятность диагноза: 0.3)

| | признаки | | | |
|---|---------------|-------------|---------|-----------------|
| | головная боль | температура | тошнота | симптом Кернига |
| % пациентов с данным признаком, болеющих менингитом | 30 | 15 | 5 | 2 |
| % пациентов с данным признаком, но без менингита | 70 | 35 | 5 | 0 |

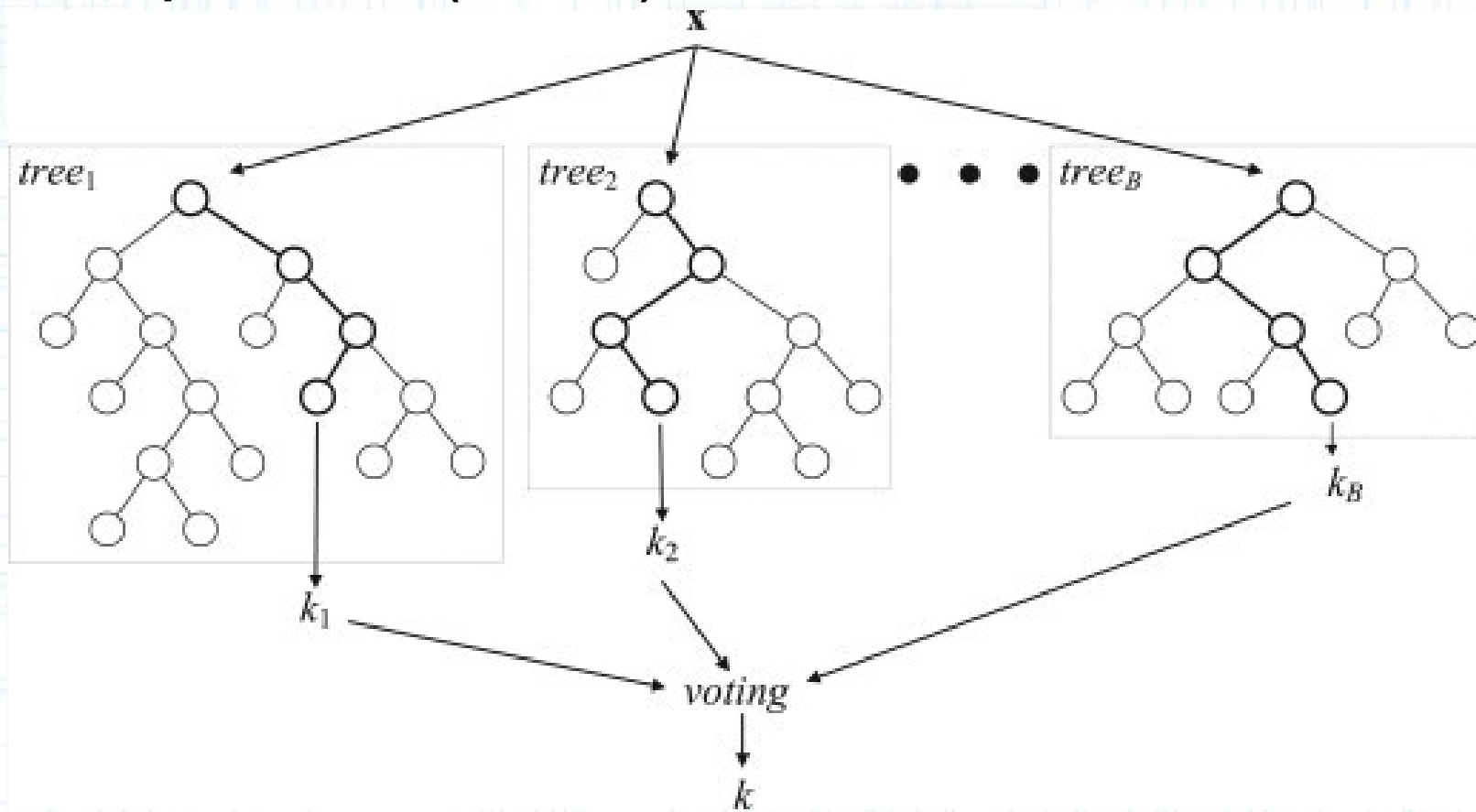
Деревья решений

- Поиск наилучших закономерностей
- Группа 100 пациентов с подозрением на менингит (вероятность диагноза: 0.3)

| | признаки | | | |
|--|---------------|-------------|---------|-----------------|
| | головная боль | температура | тошнота | симптом Кернига |
| % пациентов с данным признаком, болеющих менингитом | 30 | 15 | 5 | 2 |
| % пациентов с данным признаком, но без менингита | 70 | 35 | 5 | 0 |
| вероятность такого распределения пациентов при условии независимости признака и диагноза | 1 | 0.1725 | 0.0996 | 0.4879 |

Случайный лес

- Лео Брейман и Адель Катлер публикуют описание алгоритма, заключающегося в использовании ансамбля решающих деревьев (2002)



Развитие нейронных сетей

- 1957 – Розенблат создает простую модель перцептрона
- 1986 – метод обратного распространения ошибки
- 1998 – метод стохастического градиента
- 2000 – Игорь Айзенберг придумал термин «deep learning»
- 2009 – запуск БД изображений ImageNet
- 2011 – сверточная сеть AlexNet от Alex Krizhevsky побеждает во всех конкурсах
- 2014 – изобретение генеративных сетей

NLP – обработка естественного языка

- синтез речи
- распознавание речи
- распознавание текста (печатного и рукописного)
- машинный перевод
- информационный поиск
- ответы на вопросы

NLP

- 1988 — латентный семантический анализ (LSA)
- 2000 — латентное размещение Дирихле (LDA)
- 2000 - вероятностный латентно-семантический анализ (PLSA)
- 2013 — изобретение word2vec и начало широкого применения нейронных сетей в NLP
- 2014 — применение сетей-трансформеров для перевода
- 2016 — использование сетей на основе attention для перевода (частный случай сетей с памятью)
- 2018 — использование предобученных нейронных сетей

Компьютерные шахматы

- Deep Blue
- Deep Fritz
- Рейтинг CCRL
 - Stockfish
 - Komodo
 - Houdini
 - Fire
 - Rybka
 - Strelka



Watson



Информационный поиск

- поиск документов по запросу (разрешение лексической многозначности, морфологический анализ)
- классификация документов
- фильтрация документов
- аннотирование и реферирование документов

Яndex Google

Рекомендательные системы

- Netflix prise
- Коллаборативная фильтрация
- FastXML

The screenshot shows the Netflix website interface for a user named Jeremy Huletta. The top navigation bar includes links for 'Browse', 'Movies You'll Like', 'Friends', 'Queue', 'DVD Sale \$5.99+', 'Watch Now', and a search bar. Below the navigation bar, the main content area is titled 'Movies For You'. It features several sections of movie recommendations:

- Jeremy, here are some popular movies to get you started:** This section lists movies like 'Jackass: Number Two', 'Bill Maher: Victory Begins at Home', 'Vodka Lemon', 'Mitch Hedberg: Mitch All Together', 'It's All Gone Pete Tong', and 'Comedians of Comedy: Live at the El Rey'. Each movie entry includes a small image, a title, a star rating, and a brief description.
- OTHER MOVIES YOU MIGHT ENJOY:** This section displays a grid of movie thumbnails with titles such as 'The Film Is Not Yet Rated', 'Memoirs of a Geisha', 'March of the Penguins', 'Cinderella Man', 'Tomb Raider: The Cradle of Life', 'Serenity', 'Sweeney Todd', 'Proof', 'Lead of War', 'The Big White', 'The Weather Man', and 'North Country'.
- LOCAL FAVORITES FOR SAN FRANCISCO, CALIFORNIA:** This section lists movies like 'Prime Suspect 3 (2-Disc Series)', 'The Duchess of Duke Street Series 1 (3-Disc Series)', 'Angel Face', 'So Close', 'Chik Yeung Tin Sai', and 'Time to Leave'.

On the right side of the interface, there is a 'BROWSE' section with 'Favorite Genres' (Anime & Animation, Classics, Comedy, etc.) and 'Other Genres' (All Genres, Action & Adventure, etc.). At the bottom right, there is a promotional banner for 'Give FREE rentals!' with a 'Tell a friend' button.

Анализ тональности текста

- Тональность – позиция автора, относительно темы сообщения
- Сопряженная задача: определение темы
- Приложения:
 - политология
 - маркетинг (мнения людей о товарах, контекстная реклама)

Соционика

- построение модели личности
- методы использования показателей
- применения:
 - таргетинг рекламы
 - найм сотрудников
 - предвыборные кампании



Анализ временных рядов

за неделю за месяц за квартал за год за всё время







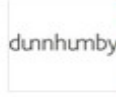

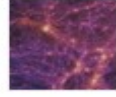


За последние 10 дней

| Дата | Курс | Изменение |
|----------|---------|-----------|
| 19.10.18 | 65,5825 | -0,2375 ↓ |
| 18.10.18 | 65,8200 | 0,3225 ↑ |
| 17.10.18 | 65,4975 | 0,1900 ↑ |
| 16.10.18 | 65,3075 | -0,3425 ↓ |
| 15.10.18 | 65,6500 | -0,4675 ↓ |
| 12.10.18 | 66,1175 | -0,1525 ↓ |
| 11.10.18 | 66,2700 | -0,6300 ↓ |
| 10.10.18 | 66,9000 | 0,6500 ↑ |
| 09.10.18 | 66,2500 | -0,3075 ↓ |
| 08.10.18 | 66,5575 | -0,0400 ↓ |

Соревнования

- [kaggle.com](https://www.kaggle.com)

| | | | | |
|---|---|----------|-----|---------------|
|  | Algorithmic Trading Challenge Develop new models to accurately predict the market response to large trades. | \$10,000 | 113 | 20 months ago |
|  | Don't Get Kicked! Predict if a car purchased at auction is a lemon | \$10,000 | 571 | 20 months ago |
|  | Give Me Some Credit Improve on the state of the art in credit scoring by predicting the probability that somebody will experience financial distress in the next two years. | \$5,000 | 926 | 21 months ago |
|  | Photo Quality Prediction Given anonymized information on thousands of photo albums, predict whether a human evaluator would mark them as 'good'. | \$5,000 | 206 | 22 months ago |
|  | Semi-Supervised Feature Learning There's been a lot of recent work done in unsupervised feature learning for classification and there are a ton of older methods that also work well. The purpose of this competition is to find out which of these methods work best on relatively large-scale high dimensional learning tasks. | \$500 | 28 | 23 months ago |
|  | Claim Prediction Challenge (Allstate) A key part of insurance is charging each customer the appropriate price for the risk they represent. Risk varies widely from customer to customer, and a deep understanding of different risk factors helps predict the likelihood and cost of insurance claims. The goal of this competition is to better predict Bodily Injury Liability Insurance claim payments based on the characteristics of the insured customer's vehicle. | \$10,000 | 105 | 23 months ago |
|  | dunnhumby's Shopper Challenge Going grocery shopping, we all have to do it, some even enjoy it, but can you predict it? dunnhumby is looking to build a model to better predict when supermarket shoppers will next visit the store and how much they will spend. | \$10,000 | 279 | 23 months ago |
|  | Wikipedia's Participation Challenge This competition challenges data-mining experts to build a predictive model that predicts the number of edits an editor will make five months from the end date of the training dataset. | \$10,000 | 94 | 24 months ago |
|  | Mapping Dark Matter Supported by NASA and the Royal Astronomical Society. A cosmological image analysis competition to measure the small distortion in galaxy images caused by dark matter. The prize is an expenses paid visit to the NASA Jet Propulsion Laboratory (JPL). | \$3,000 | 72 | 2 years ago |

Репозиторий UCI

- Ссылка

UCI



Machine Learning Repository

Center for Machine Learning and Intelligent Systems

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Repository Web

Google™

[View ALL Data Sets](#)

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 481 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#).

Supported By:



In Collaboration With:



Latest News:


- 09-24-2018:** Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!
- 04-04-2013:** Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
- 03-01-2010:** [Note](#) from donor regarding Netflix data
- 10-16-2009:** Two new data sets have been added.
- 09-14-2009:** Several data sets have been added.
- 03-24-2008:** New data sets have been added!
- 06-25-2007:** Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: [Abalone](#)




Task: Classification
Data Type: Multivariate
Attributes: 8
Instances: 4177

Newest Data Sets:

- 07-30-2019:**  [PPG-DaLiA](#)
- 07-24-2019:**  [Divorce Predictors data set](#)
- 07-22-2019:**  [Alcohol QCM Sensor Dataset](#)
- 07-14-2019:**  [Incident management process enriched event log](#)
- 06-30-2019:**  [Wave Energy Converters](#)
- 06-22-2019:**  [Query Analytics Workloads Dataset](#)

Most Popular Data Sets (hits since 2007):

- 2808763:**  [Iris](#)
- 1568695:**  [Adult](#)
- 1217307:**  [Wine](#)
- 1028814:**  [Car Evaluation](#)
- 1009132:**  [Wine Quality](#)
- 998034:**  [Heart Disease](#)