

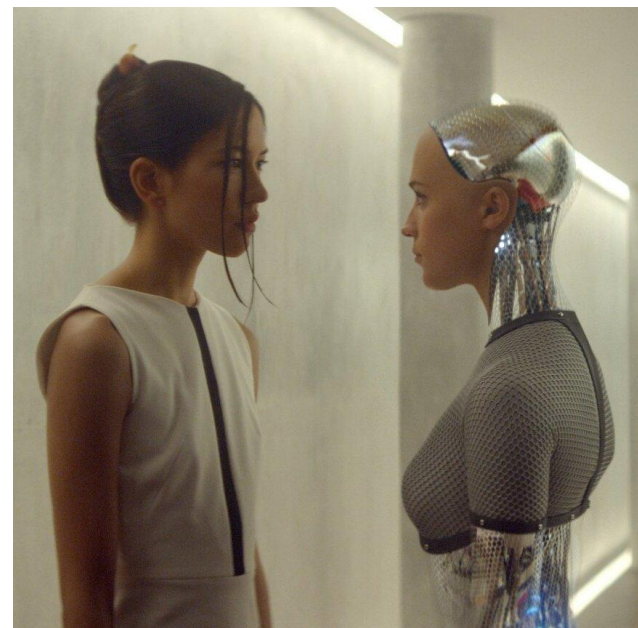
Машинное обучение

История и методология прикладной
математики и информатики

Ю.Б.Буркатовская, доцент ОИТ

Что это?

- **Машинное обучение** (англ. **machine learning**, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а **обучение** в процессе применения решений множества сходных задач.
- **Искусственный интеллект** (ИИ; англ. **artificial intelligence**, AI) — свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека; наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ.



Что это?

- **Машинное обучение — это...**
- Машинное обучение — это «[...] класс методов искусственного интеллекта, которые позволяют улучшить результаты работы компьютеров путем обучения на известных данных», — Berkeley.
- **...подраздел искусственного интеллекта (ИИ)**
- ИИ — это наука и технология по разработке мероприятий и методов, позволяющих компьютерам успешно выполнять задачи, которые обычно требуют интеллектуального осмысления человека. Машинное обучение — часть этого процесса: это методы и технологии, с помощью которых можно обучить компьютер выполнять поставленные задачи.
- **...способ решения практических задач**
- Методы машинного обучения все еще в развитии. Некоторые уже изучены и используются, но ожидается, что со временем их количество будет только расти. Идея в том, что совершенно разные методы используются для совершенно разных компьютеров, а различные бизнес-задачи требуют различных методов машинного обучения.

Что это?

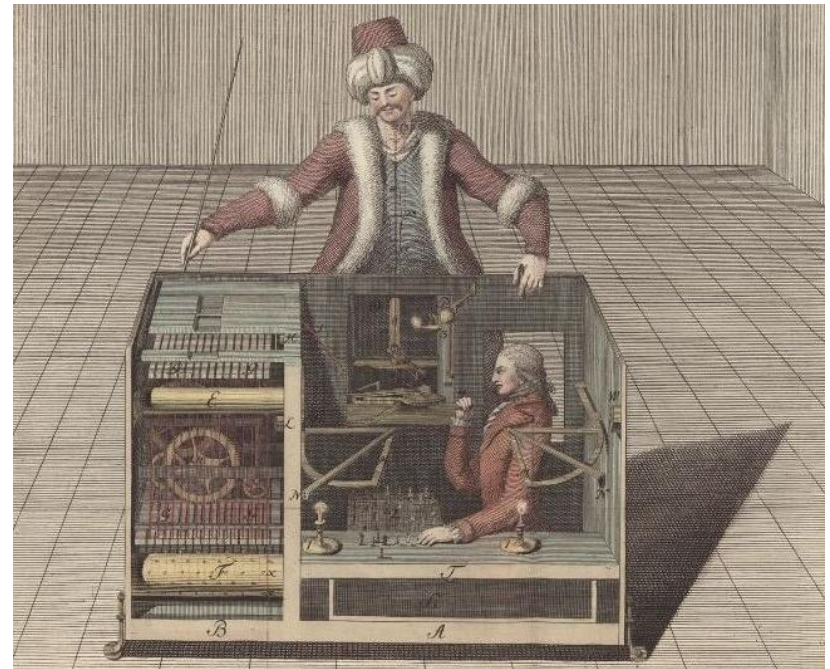
- **... способ увеличить эффективность компьютеров**
- Для решения компьютером задач с применением искусственного интеллекта нужны практика и автоматическая поднастройка. Модель машинного обучения нуждается в тренировке с использованием базы данных и в большинстве ситуаций — в подсказке человека.
- **...технология, основанная на опыте**
- ИИ нуждается в предоставлении опыта — иными словами, ему необходимы данные. Чем больше в систему ИИ поступает данных, тем точнее компьютер взаимодействует с ними, а также с теми данными, что получает в дальнейшем. Чем выше точность взаимодействия, тем успешнее будет выполнение поставленной задачи, и выше степень прогностической точности.
- <https://netology.ru/blog/machine-learning-guide>

Что это?



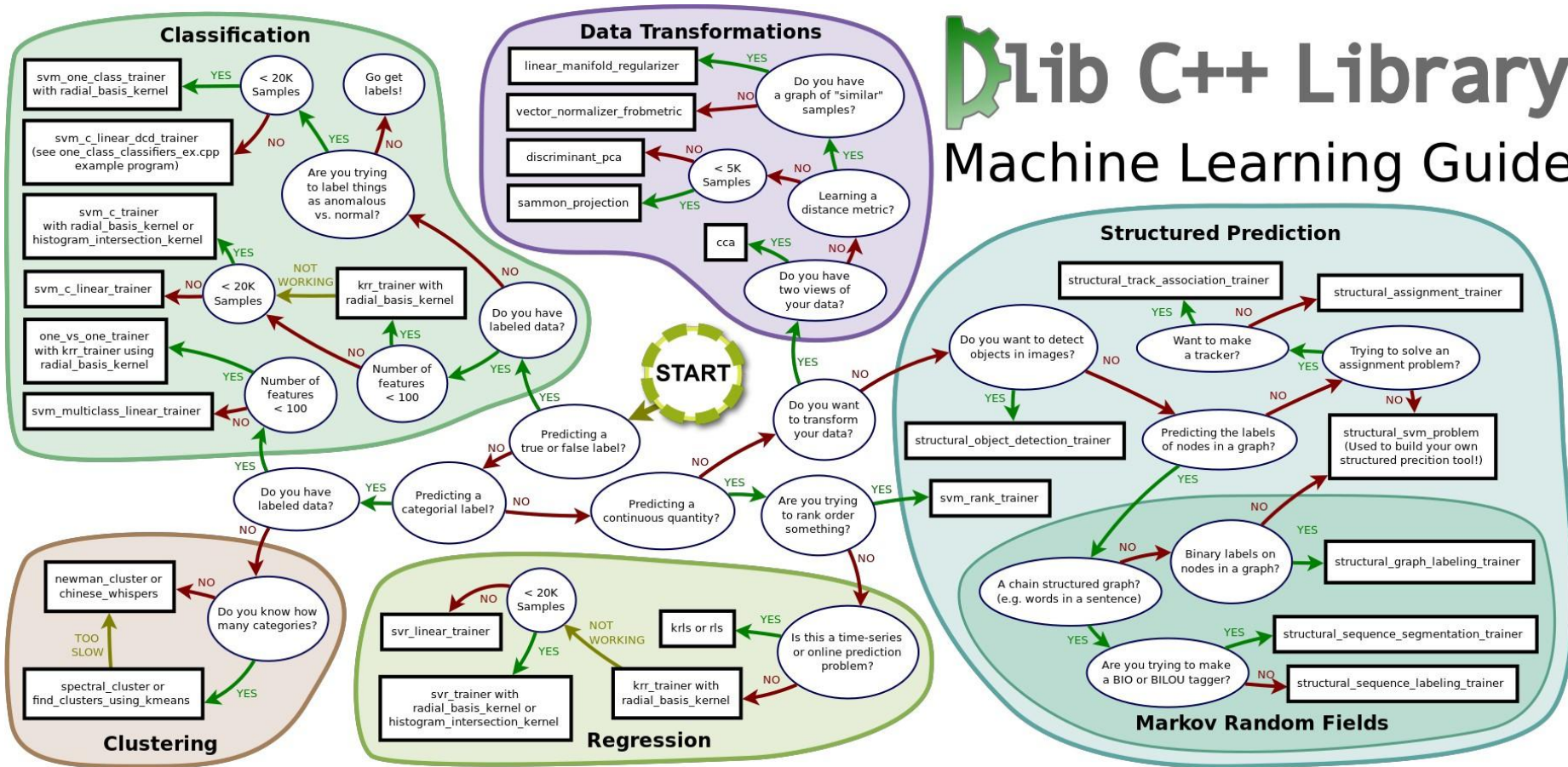
Что это?

- **1770 г.** Появляется первый шахматный автомат, который дурачил Европу десятки лет. Механическое устройство под названием «Турок» обмануло самого Наполеона. Внутри прятался сильный шахматист и управлял шахматным автоматом. Со стороны казалось, что это чудомашина, которая наделена разумом. Обман раскрылся только в 1857 году, когда последний владелец устройства показал, где прятался человек и как он управлял «Турком».



Что это?

Lib C++ Library Machine Learning Guide



Что это?

Задачи

- Классификация.
- Кластеризация.
- Регрессия.
- Понижение размерности данных.
- Восстановление плотности распределения.
- Одноклассовая классификация и выявление новизны.
- Построение ранговых зависимостей.
- Поиск ассоциативных правил
- Фильтрация выбросов.
- Заполнение пропущенных значений.

База

- Теория вероятностей
- Математическая статистика
- Методы оптимизации
- Линейная алгебра
- Математическая логика

Что это?

Типы

- **Обучение с учителем** (supervised learning)
- **Обучение без учителя, или самообучение** (unsupervised learning).
- **Обучение с подкреплением** (reinforcement learning)
- **Частичное обучение** (semi-supervised learning)
- **Динамическое обучение** (online learning)
- **Активное обучение** (active learning)

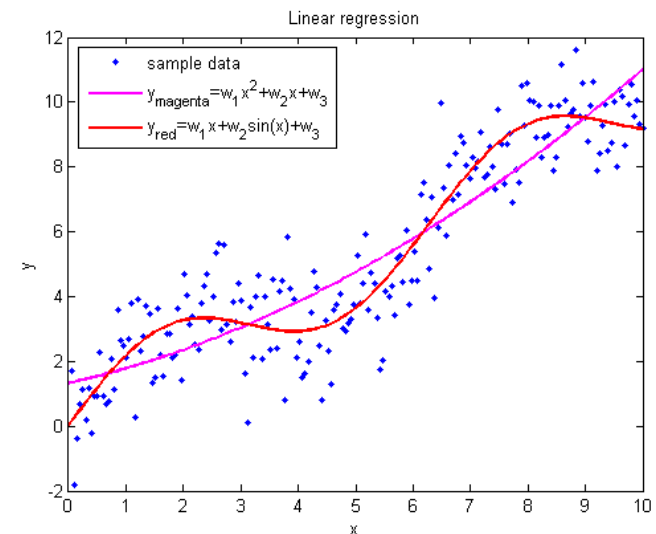
Методы

- **Статистические**
 - Байесовские методы
 - Деревья решений
 - KNN (к ближайших соседей)
 - SVM (опорная машина векторов)
 - PCA (метод главных компонент)
 - K-means (к средних)
- **Обучение с подкреплением**
 - Q-learning
 - Генетические алгоритмы
- **Ансамблевые**
 - Бустинг
 - Стекинг
 - Бэггинг
- **Нейросети** (перцептрон, CNN, RNN...)

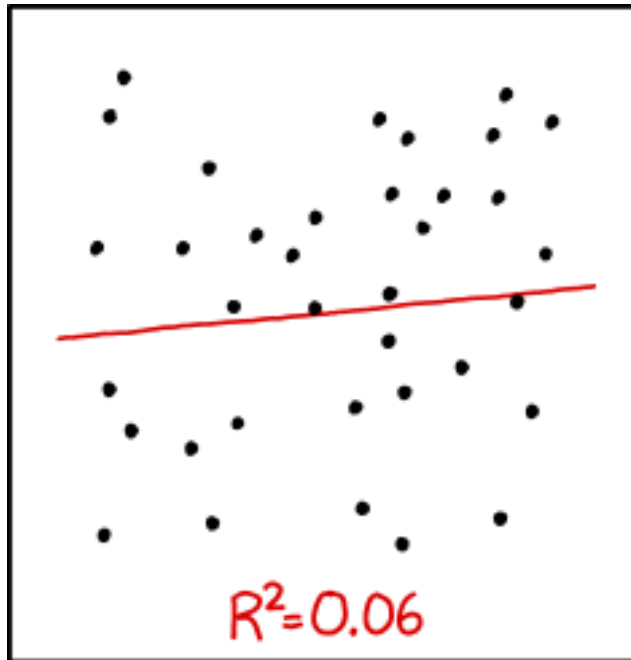
Задачи ML

Регрессия (обучение с учителем).

- Данные состоят из пар значений переменных: **зависимой** (переменной отклика) и **независимой** (объясняющей переменной).
- **Регрессионная модель** есть функция независимой переменной и параметров с добавленной случайной переменной.
- Параметры модели настраиваются таким образом, что модель наилучшим образом приближает данные.
- Регрессионный анализ используется для прогноза, анализа временных рядов, проверки гипотез и выявления скрытых взаимосвязей в данных.



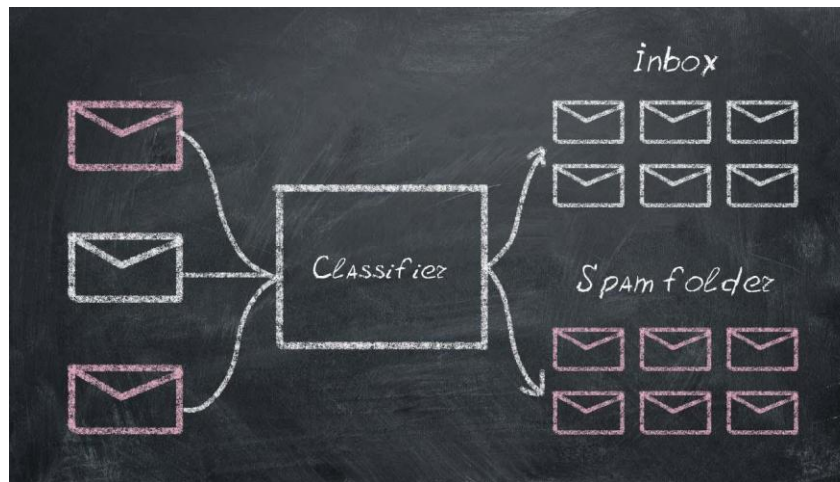
Задачи ML



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Задачи ML

- **Классификация (обучение с учителем).**
- Имеется множество *объектов* (ситуаций), разделённых некоторым образом на *классы*. Задано конечное множество объектов, для которых известно, к каким классам они относятся (*обучающая выборка*).
- Требуется классифицировать (отнести к одному из классов) произвольный объект из исходного множества.
- Классификацию сигналов и изображений называют также *распознаванием образов*.



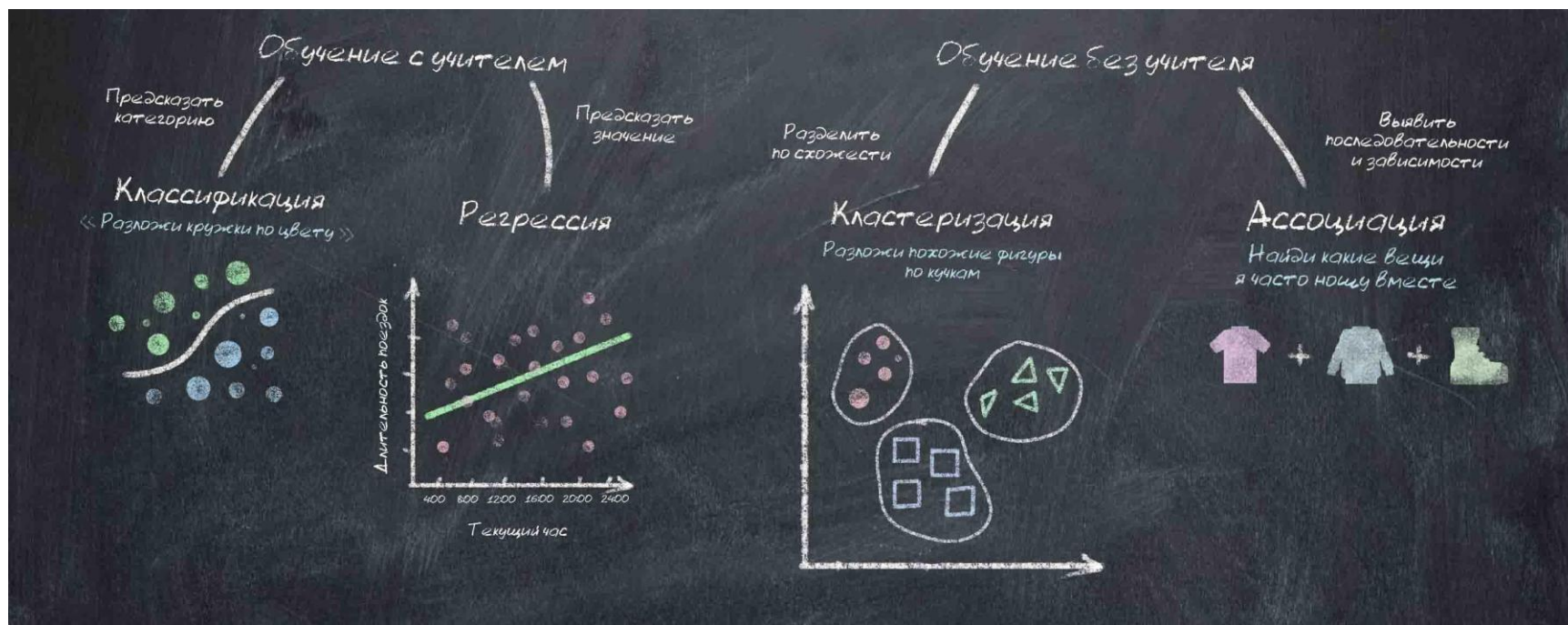
Задачи ML

- **Прогнозирование (обучение с учителем).**
- Под прогнозированием понимают предсказание будущих значений временного ряда на основе его значений в прошлом, и, возможно, дополнительной информации. Такую дополнительную информацию представляют влияющие на ситуацию внешние факторы. Например, для прогнозирования временного ряда спроса на какой-либо товар это могут быть предложения замещающих товаров, для экспортных грузоперевозок — таможенные пошлины и курс доллара, для цен акций — политические решения и т. д.
- **Ранжирование (обучение с учителем).**
- Задача **ранжирования** (learning to rank) отличается тем, что ответы надо получить сразу на множестве объектов, после чего отсортировать их по значениям ответов. Может сводиться к задачам классификации или регрессии. Часто применяется в информационном поиске и анализе текстов.

Задачи ML

- **Обучение без учителя.**
- **Кластеризация.**
- Задача разбиения заданной выборки *объектов* (ситуаций) на непересекающиеся подмножества, называемые *кластерами*, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.
- Каждый объект описывается набором своих характеристик, называемых *признаками*. Признаки могут быть числовыми или нечисловыми.
- Каждый объект описывается расстояниями до всех остальных объектов обучающей выборки.
- **Поиск ассоциативных правил.**
- «Кто купил x, также купил y». В основе лежит анализ транзакций, каждая из которых представляет собой уникальный набор товаров. При помощи ARL алгоритмов находятся те самые «правила» сочетания товаров внутри одной транзакции, которые потом сортируются.

Задачи ML



Задачи ML

- **Обучение без учителя.**
- Задача **фильтрации выбросов** (outliers detection) — обнаружение в обучающей выборке небольшого числа нетипичных объектов. В некоторых приложениях их поиск является самоцелью (например, обнаружение мошенничества). В других приложениях эти объекты являются следствием ошибок в данных или неточности модели, то есть шумом, мешающим настраивать модель, и должны быть удалены из выборки.
- Задача **сокращения размерности** (dimensionality reduction) заключается в том, чтобы по исходным признакам с помощью некоторых функций преобразования перейти к наименьшему числу новых признаков, не потеряв при этом никакой существенной информации об объектах выборки.
- Задача **заполнения пропущенных значений** (missing values) — замена недостающих значений в матрице объекты–признаки их прогнозными значениями.

Задачи ML

- **Обучение с подкреплением** (reinforcement learning). Роль объектов играют пары «ситуация, принятое решение», ответами являются значения функционала качества, характеризующего правильность принятых решений (реакцию среды). Как и в задачах прогнозирования, здесь существенную роль играет фактор времени. Примеры прикладных задач: формирование инвестиционных стратегий, автоматическое управление технологическими процессами, самообучение роботов, и т.д.
- **Динамическое обучение** (online learning) может быть как обучением с учителем, так и без учителя. Специфика в том, что прецеденты поступают потоком. Требуется немедленно принимать решение по каждому прецеденту и одновременно доучивать модель зависимости с учётом новых прецедентов.
- **Активное обучение** (active learning) отличается тем, что обучаемый имеет возможность самостоятельно назначать следующий прецедент, который станет известен (планирование эксперимента).

Задачи ML

- **Частичное обучение** (semi-supervised learning) занимает промежуточное положение между обучением с учителем и без учителя. Каждый прецедент представляет собой пару «объект, ответ», но ответы известны только на части прецедентов. Пример прикладной задачи — автоматическая рубрикация большого количества текстов при условии, что некоторые из них уже отнесены к каким-то рубрикам.
- **Трансдуктивное обучение** (transductive learning). Дана конечная обучающая выборка прецедентов. Требуется по этим *частным* данным сделать предсказания относительно других *частных* данных — тестовой выборки. В отличие от стандартной постановки, здесь не требуется выявлять *общую* закономерность, поскольку известно, что новых тестовых прецедентов не будет. С другой стороны, появляется возможность улучшить качество предсказаний за счёт анализа всей тестовой выборки целиком, например, путём её кластеризации. Во многих приложениях трансдуктивное обучение практически не отличается от частичного обучения.

Основные открытия до ML

- **1763.** Статья **Томаса Байеса** «*An Essay towards solving a Problem in the Doctrine of Chances*» публикуется через два года после его смерти, с поправками и изменениями, внесенными другом Байеса Ричардом Прайсом.
- **1805.** **Адриен-Мари Лежандр** описывает **метод наименьших квадратов**. Применялся Гауссом и Лапласом.
- **1812.** **Пьер-Симон Лаплас** публикует «*Théorie Analytique des Probabilités*», в которой он расширяет работу Байеса и определяет то, что сейчас известно как **теорема Байеса**.
- **1886.** **Френсис Гальтон** предлагает **регрессионную модель**.
- **1901.** **Метод главных компонент** изобретен **Карлом Пирсоном**.
- **1913.** **Андрей Марков** впервые описывает приемы, которые он использовал для анализа «Евгения Онегина», ставшие основой теории **цепей Маркова**.
- **1950.** **Алан Тьюринг** в статье «*Computing Machinery and Intelligence*» предложил **тест Тьюринга**. Компьютер «мыслит», если человек, взаимодействующий с ним, не сможет в процессе общения отличить компьютер от другого человека.

История ML

- **1943.** Уолтер Питтс и Уоррен МакКаллох в статье “*A Logical Calculus of the Ideas Immanent in Nervous Activity*” («Логическое исчисление идей, относящихся к нервной деятельности») показывают математическую модель биологического нейрона.
- Этот нейрон МакКаллока Питтса имеет очень ограниченные возможности и не имеет механизма обучения. Тем не менее, это заложит основу для искусственной нейронной сети и глубокого обучения.
- Основная идея – описание нервной деятельности языком логики высказываний.

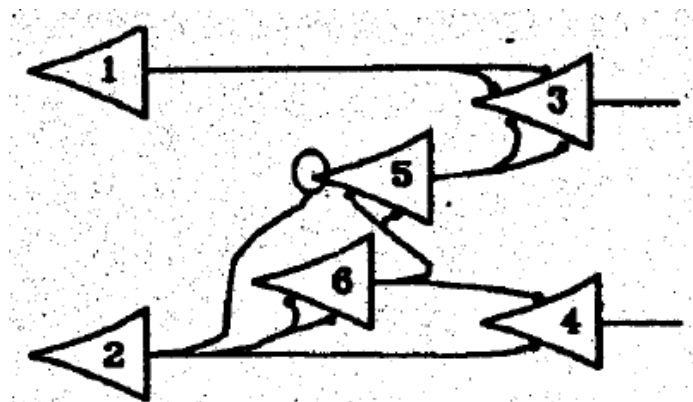
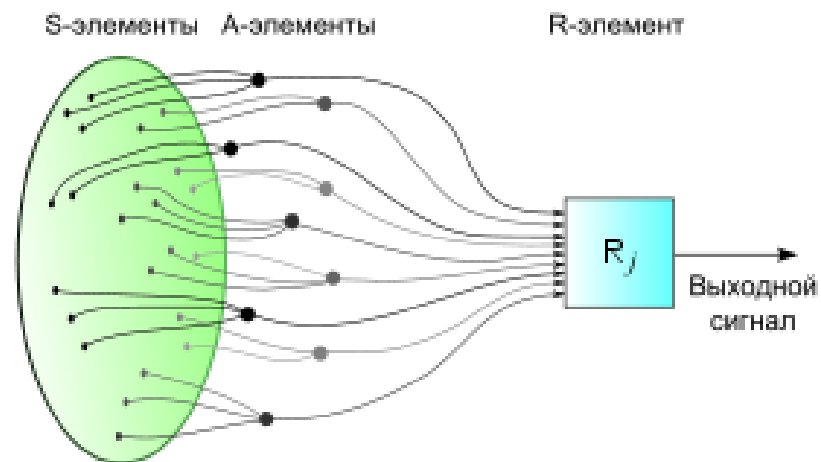


Рис. 5.
 $N_3(t) = N_1(t-1) \cdot V$
 $N_2(t-3) = N_2(t-2)$,
 $N_4(t) = N_2(t-2) \cdot N_2(t-1)$.

История ML

- **1949.** Д. Хебб в работе «Организация поведения» описал основные принципы обучения нейронов.
- **1951.** Американский нейрофизиолог **Фрэнк Розенблатт** в статье “*The Perceptron: A Perceiving and Recognizing Automaton*” («Перцептрон: воспринимающий и распознающий автомат») демонстрирует перцептрон, который имел истинные способности к обучению для самостоятельной бинарной классификации.



История ML

- Элементарный перцептрон состоит из элементов трёх типов: S-элементов, A-элементов и *одного* R-элемента.
- S-элементы — это слой сенсоров или рецепторов. В физическом воплощении они соответствуют, например, светочувствительным клеткам сетчатки глаза или фоторезисторам матрицы камеры. Каждый рецептор может находиться в состоянии *покоя* или *возбуждения*, и только в последнем случае он передаёт единичный сигнал в следующий слой, ассоциативным элементам.
- A-элементы называются ассоциативными, потому что каждому такому элементу, как правило, соответствует целый набор (ассоциация) S-элементов. A-элемент активизируется, как только количество сигналов от S-элементов на его входе превысило некоторую величину θ .
- Сигналы от возбуждёвшихся A-элементов, в свою очередь, передаются в сумматор R, причём сигнал от i -го ассоциативного элемента передаётся с коэффициентом w_i . Этот коэффициент называется *весом* A—R связи.

История ML

- R-элемент подсчитывает сумму значений входных сигналов, помноженных на веса. R-элемент выдаёт «1», если сумма превышает порог θ , иначе на выходе будет «-1».
- Обучение элементарного перцептрона состоит в изменении весовых коэффициентов связей A—R. Веса связей S—A (которые могут принимать значения $\{-1; 0; +1\}$) и значения порогов A-элементов выбираются случайным образом в самом начале и затем не изменяются.
- **Метод коррекции ошибки** — вес связи не изменяется до тех пор, пока текущая реакция перцептрона остается правильной. При появлении неправильной реакции вес изменяется на единицу, а знак (+/-) определяется противоположным от знака ошибки.
- Далее перцептрон готов работать в режиме *распознавания* или *обобщения*. В этом режиме перцептрону предъявляются ранее неизвестные ему объекты, и перцептрон должен установить, к какому классу они принадлежат.

История ML

- Согласно современной терминологии, перцептроны – это *искусственные нейронные сети*:
 - с одним скрытым слоем;
 - с пороговой передаточной функцией;
 - с прямым распространением сигнала.

История ML

Персептрон

Модель восприятия информации мозгом, состоящая из S, A и R элементов

Персептрон с одним
скрытым слоем

1 слой S элементов
1 слой A элементов
1 слой R элементов

Однослойный
персептрон

Каждому S только
один A
S-A Связи = +1
Порог A = +1

Многослойный
персептрон по
Розенблатту

> 1 слоя A элем.

Многослойный
персептрон по
Румельхарту

S-A Обучаются
Спец алг. обуч.

История ML

- **1952.** Марвин Мински и Дин Эдмондс строят первый нейронный сетевой компьютер, способный к обучению, SNARC (Stochastic Neural Analog Reinforcement Calculator).
- Сама машина представляет собой случайно подключенную сеть из примерно 40 синапсов Хебба. Каждый из этих синапсов имеет память, в которой хранится вероятность того, что сигнал поступит на вход, а другой сигнал появится на выходе. Есть “ручка вероятности”, которая может находиться в положении от 0 до 1 и показывает вероятность распространения сигналов. Если сигнал проходит, конденсатор запоминает эту функцию и включает «сцепление». В этот момент оператор нажимает кнопку, чтобы дать вознаграждение машине. Запускается большой двигатель, и на всех 40 синапсах проверяется, включено ли сцепление. Поскольку конденсатор может «запоминать» только определенное время, запоминаются только самые последние обновления вероятностей.

История ML

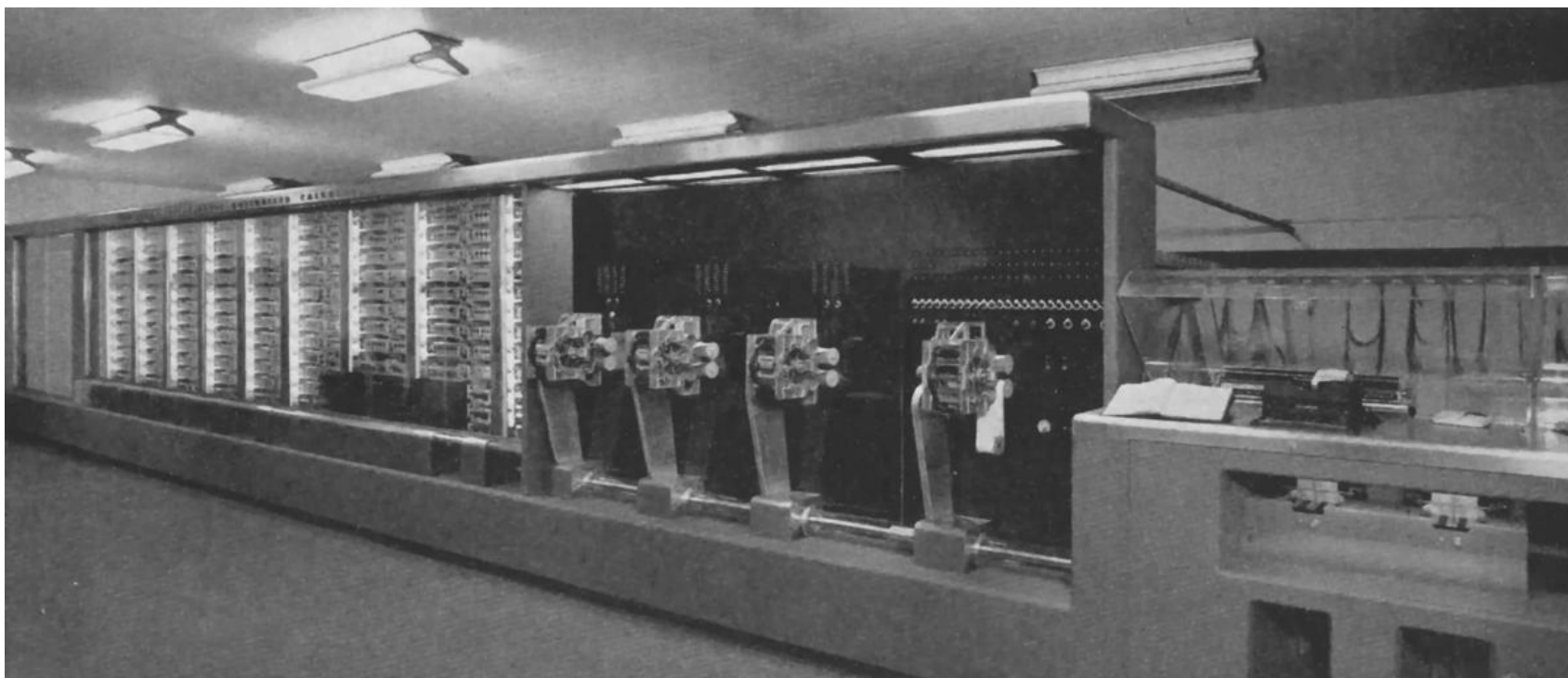
- **1952.** Артур Самуэль в лаборатории IBM создает программы Checkers-playing , которые играют в шашки.
- Основой программного механизма было **дерево поиска** игровых позиций, достижимых из текущего состояния. Так как количество памяти, бывшее в его распоряжении, было ограниченным, Самуэль реализовал то, что теперь называют **альфа бета отсечением**. Вместо того, чтобы искать каждый путь для всего хода игры, Самуэль развивал сложную полиномиальную оценочную функцию, основанную на позиции в любой момент времени.
- Программа выбирала ход, основанный на **минимаксной** стратегии, то есть она делала ход, который максимизировал ценность позиции, предполагая, что противник делает то же самое со своей стороны.
- Более поздние программы Самуэля переоценивали веса оценочной функции, на основе игр профессионалов. Также он заставлял программу играть против самой себя и таким образом самообучаться.

История ML

- В **1957** году в Корнелльской Лаборатории Аэронавтики успешно было завершено моделирование работы перцептрона на компьютере IBM 704, а два года спустя, 23 июня 1960 года в Корнелльском университете, был продемонстрирован первый нейрокомпьютер — **«Марк-1»**, который был способен распознавать некоторые из букв английского алфавита.

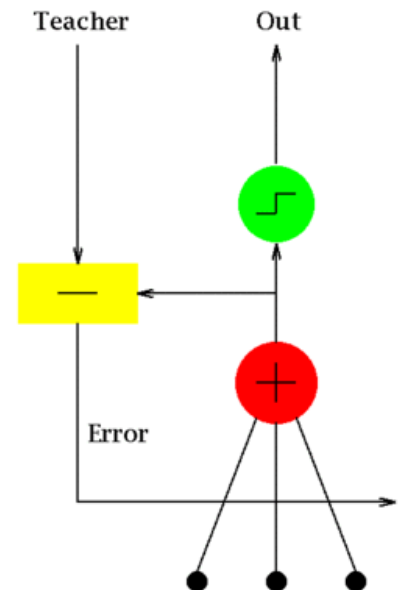


История ML



История ML

- **1959 г. Искусственная нейронная сеть научилась удалять помехи во время телефонных разговоров.** Первой нейронной сетью, которую использовали для решения реальной проблемы, стала MADALINE (Adaptive Linear Neuron или более поздний Adaptive Linear Element). Она с помощью адаптивного фильтра удаляла эхо-сигналы на телефонных линиях.
- Он был разработан профессором Бернардом Уидроу и его аспирантом Тедом Хоффом в Стэнфордском университете.
- Разница между Adaline и стандартным персептроном заключается в том, что на этапе обучения веса корректируются в соответствии с взвешенной суммой входных данных. В стандартном персептроне для регулировки весов используется выход функции.



История ML

- **1960.** Генри Дж. Келли в статье *«Градиентная теория оптимальных траекторий полета»* демонстрирует первую в истории версию модели непрерывного обратного распространения (**backpropagation**). Его модель находится в контексте теории управления, но она закладывает основу для дальнейшего совершенствования модели и будет использоваться в нейросетях в будущем.
- **1962.** Стюарт Дрейфус в статье *«Численное решение вариационных задач»* показывает модель обратного распространения, которая использует простое правило цепочки производных вместо динамического программирования, которое использовались в более ранних моделях обратного распространения. Это еще один маленький шаг, который укрепляет будущее глубокого обучения.

История ML

- **1962. Развитие идеи перцептрона.**
- Ф. Розенблатт. *«Принципы нейродинамики: Перцептроны и теория механизмов мозга»*. В книге рассматриваются не только уже готовые модели перцептрона с одним скрытым слоем, но и многослойных перцептронов с перекрёстными и обратными (четвёртая глава) связями. В книге также вводится ряд важных идей и теорем, например, доказывается теорема сходимости перцептрона.
- *Элементарный перцептрон, обучаемый по методу коррекции ошибки (с квантованием или без него), независимо от начального состояния весовых коэффициентов и последовательности появления стимулов всегда приведет к достижению решения за конечный промежуток времени.*

История ML

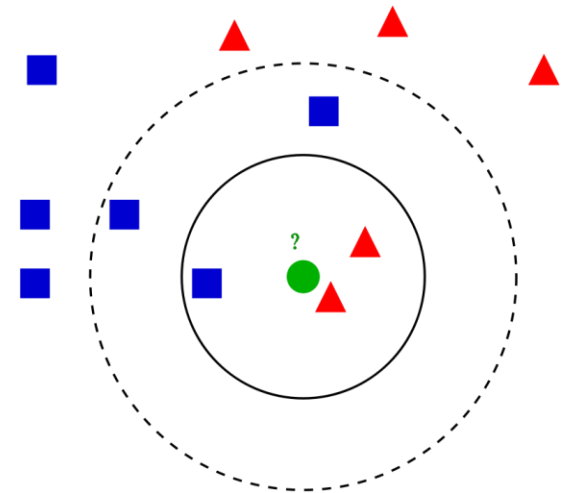
- **1963.** Дональд Мичи создает «машину», состоящую из 304 спичечных коробок и бусин, которая использует обучение с подкреплением для игры в крестики-нолики. Называется эта машина MENACE (Mathbox Educable Naughts and Crosses Engine — машина из спичечных коробков, умеющая играть в крестики и нолики; menace (англ.) — угроза, опасность.))
- Первый двухдневный турнир между Мичи и его машиной состоял из 220 партий. Сначала Мичи все время наказывал свое детище за плохую игру, но после семнадцати партий машина начала ставить первый крестик только в угловую клетку, а после двадцатой партии заканчивать все игры вничью. В надежде заманить противника в ловушку Мичи начал делать самые бессмысленные ходы. Такая тактика оправдывала себя лишь до тех пор, пока машина не научилась справляться и с этими хитростями. Закончился матч поражением Мичи: он выбыл из турнира, проиграв восемь партий из десяти.
- **1963.** Ларри Робертс сформулировал тезисы компьютерного зрения в своей диссертации в MIT.

История ML

- **1965. Рождение многослойной нейронной сети.**
- **Алексей Григорьевич Ивахненко** вместе с **Валентином Григорьевичем Лапой** создает иерархическое представление нейронной сети, которое использует функцию полиномиальной активации и обучается с использованием **метода группового учета аргументов (МГУА)**. В настоящее время эта сеть считается первым в мире многослойным персептроном, а Ивахненко часто считается отцом глубокого обучения.
- Метод МГУА основан на рекурсивном селективном отборе моделей, на основе которых строятся более сложные модели. Точность моделирования на каждом следующем шаге рекурсии увеличивается за счет усложнения модели. В качестве моделей часто берутся полиномы, поскольку любую непрерывную на конечном интервале функцию можно со сколь угодно высокой точностью представить в виде полинома определенной степени.
- *А. Г. Ивахненко, В. Г. Лапа. Кибернетические предсказывающие устройства. — К.: «Наук. думка», 1965.*

История ML

- **1967.** Метод **k-ближайших соседей** (*k-nearest neighbors algorithm*, k-NN) — метрический алгоритм для автоматической **классификации** объектов или **регрессии**.
- В случае классификации объект присваивается тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны.
- В случае использования метода для регрессии объекту присваивается среднее значение по k ближайшим к нему объектам, значения которых уже известны.
- Алгоритм может быть применим к выборкам с большим количеством атрибутов (многомерным).



История ML

- **1967. Метод k-ближайших соседей**
- Ключевой момент метода – определение **функции расстояния** между объектами. В многомерном евклидовом пространстве можно использовать евклидову метрику, расстояние нью-йоркского таксиста, и т.д.
- **Проблема** – что если объект описывается признаками различной природы (номинальные, количественные, порядковые)?
- **Пример** – медицинские данные (результаты анализов, характер боли, кардиограмма...)

История ML

- **1969. Отказ от коннективизма. «Зима» нейросетей.**
- Исследования по ИИ в 50-е годы XX века разделились на два направления: коннективизм и символизм.
- **Коннективизм** (иногда его называют «восходящим» методом) предполагает кибернетический или нейромодельный подход к ИИ — путь от простых аналогов нервной системы человека, примитивных биологических организаций с малым числом нейронов к более сложным (перцептрон, SNARC).
- **Символизм** (его называют «нисходящим» методом) заключается в создании компьютерных программ для решения задач, требующих от человека значительного интеллекта (доказательство теорем, игра в шахматы и т. п.)
- Символизм начал побеждать после успеха программ логического вывода (Logic Theorist, General Problem Solver.)

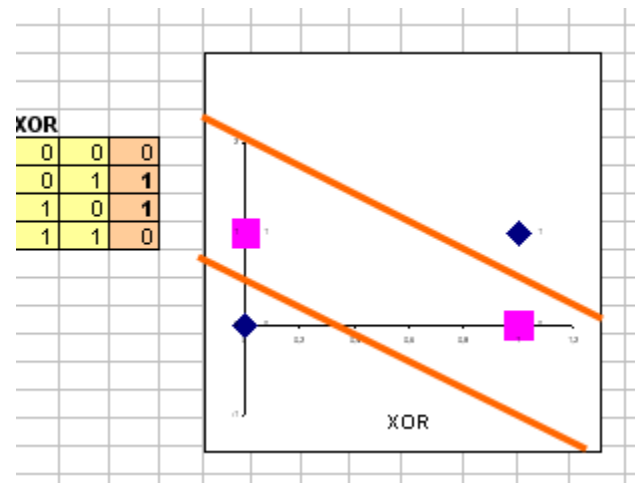
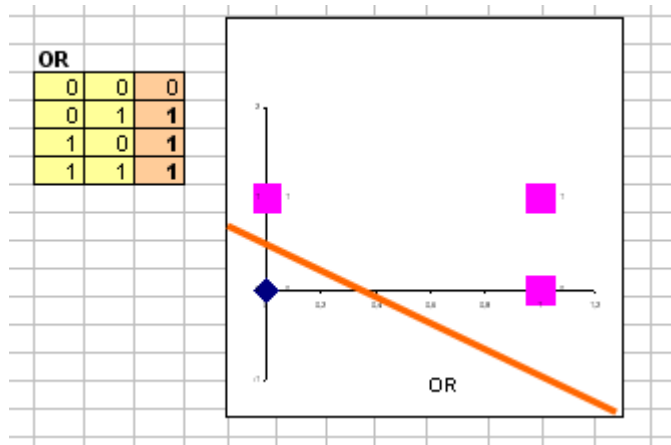
История ML

- **1969. Отказ от коннективизма. «Зима» нейросетей.**
- Однако работа в одном из направлений коннекционизма продолжалась — изучение перцептрона, предложенного Фрэнком Розенблаттом, которому удавалось поддерживать исследования в этой области благодаря своим способностям «продавца» и силы своей личности. Он оптимистично прогнозировал, что перцептрон *«со временем сможет учиться, принимать решения и переводить с одного языка на другой»*.
- **Марвин Мински и Сеймур Паперт** публикуют книгу «Перцептроны», в которой они показывают, что перцептрон Розенблатта не может обучиться вычислять сложные функции, такие как XOR. Для такой функции перцептроны должны быть размещены в нескольких скрытых слоях, что ставит под угрозу алгоритм обучения перцептрона.
- *Они никогда не обретут даже умения распознавать предмет, частично заслоненный другими. Глядя на торчащий из-за кресла кошачий хвост, подобная машина никогда не сможет понять, что она видит.*

История ML

AND (разделяется прямой)

XOR (не разделяется)



История ML

- **1969. Отказ от коннективизма. «Зима» нейросетей.**
- Авторитет Минского был столь велик, что правительственные организации прекратили финансирование работ коннективистского направления. И наоборот – деньги были переадресованы на исследования символистского направления (хотя сам Минский выразил сожаление по поводу случившегося). Хотя важная работа в этом направлении частично продолжалась, например, был предложен **метод обратного распространения**, однако найти достаточное финансирование для проектов коннекционистов в 1970-х и начале 1980-х годов было трудно.
- «Зима» исследований коннекционистов прекратилась в середине 1980-х годов, когда работы **Джона Хопфилда, Дэвида Румельхарта** и других возродили масштабный интерес к нейронным сетям.
- Розенблатт не дождался этого, он погиб в результате несчастного случая вскоре после публикации книги «Перцептроны», он погиб в свой 43 день рождения при крушении лодки.

История ML

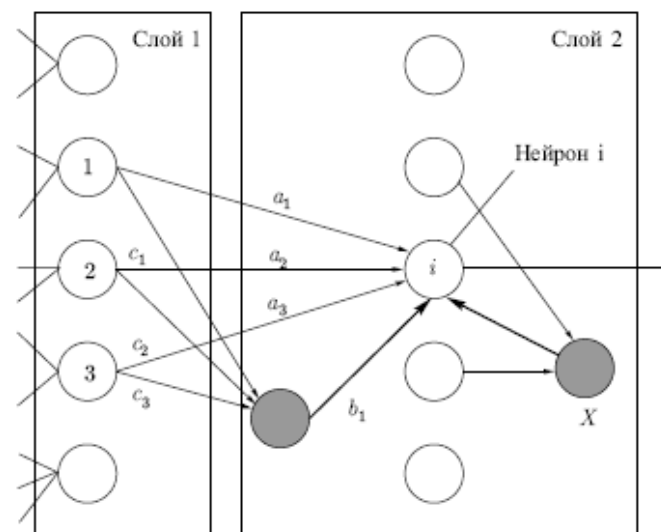
- **1970. Программная реализация backpropagation**
- **Сеппо Линмаинмаа** публикует общий метод автоматического дифференцирования для обратного распространения, а также программно реализует метод.
- Исследование обратного распространения в это время зашло очень далеко, но оно не будет применено в нейронной сети до следующего десятилетия.
- **1971. Нейронная сеть становится «глубже».**
- **Алексей Григорьевич Ивахненко** продолжает исследования в области нейронных сетей. Он создает 8-слойную глубокую нейронную сеть, используя метод группового учета аргументов.

История ML

- **1974. Обратное распространение предложено использовать в нейросетях.**
- **Пол Вербос** в докторской диссертации предлагает использование обратного распространения для распространения ошибок во время обучения нейронных сетей. Его результаты в конечном итоге приведут к практическому использованию обратного распространения в нейронных сетях в будущем.
- Независимо и одновременно **Александр Иванович Галушкин** также предложил использовать обратное распространение в нейросети.
- Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы.

История ML

- **1975. Когнитрон (Кунихико Фукушима)**
- Когнитрон состоит из иерархически связанных слоев нейронов двух типов — тормозящих и возбуждающих. Состояние возбуждения каждого нейрона определяется суммой его тормозящих и возбуждающих входов. Синаптические связи идут от нейронов одного слоя (далее слоя 1) к следующему (слою 2).
- Пробраз сети был позаимствован из модели, предложенной Хьюбелом и Визелем (1959 г.), согласно которой, существует два вида клеток в первичной зрительной коре: простая и сложная клетка, расположенные каскадно.



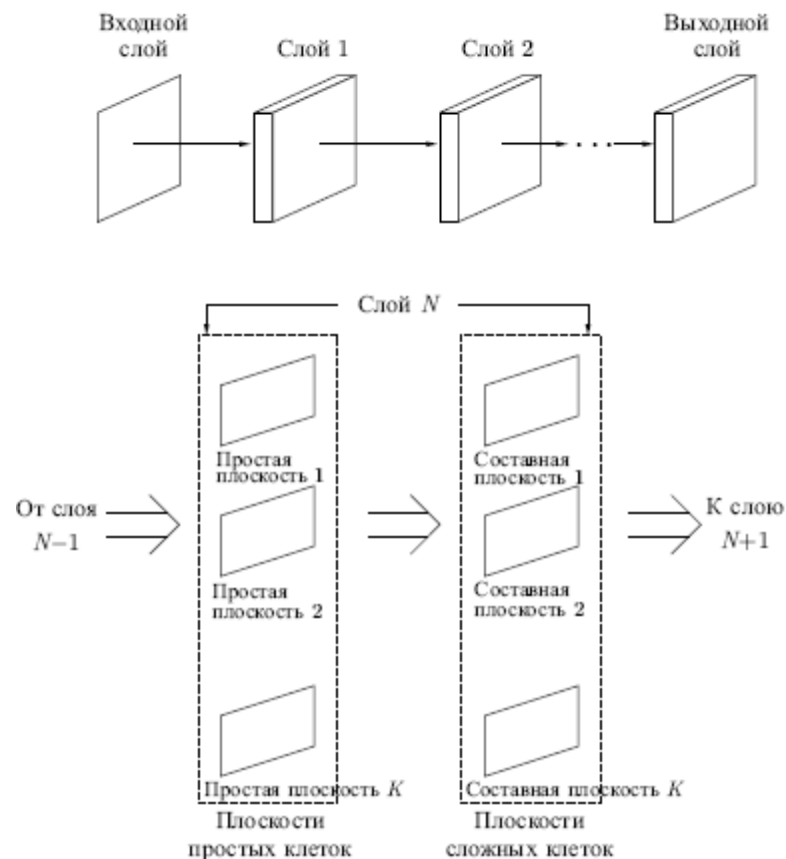
История ML

- **1979. Стэнфордская тележка.**
- Разрабатывалась как прототип беспилотного вездехода, например, для исследований на Луне и Марсе.
- Первый образец машины (1961) был связан с компьютерным "мозгом" по кабелю.
- К 1966 году машина научилась уверенно "держаться разметку" и без остановок кататься "по белой линии".
- К 1977 году "интеллект" удалось, наконец, погрузить на саму тележку, а в 1979 году с помощью процессора KL10 с производительностью 2,5 MIPS и стереоскопического "зрения" была достигнута "глубина планирования" до 20 метров, при трех объектах в поле зрения.



История ML

- **1980. Неокогнитрон** - первая архитектура CNN (сверточных нейросетей)
- Кунихико Фукушима предлагает неокогнитрон, первую сверточную архитектуру нейронных сетей, которая может распознавать визуальные шаблоны, такие как рукописные символы.
- Неокогнитрон является намного более мощным средством распознавания образов независимо от их преобразований, вращений, искажений и изменений масштаба.

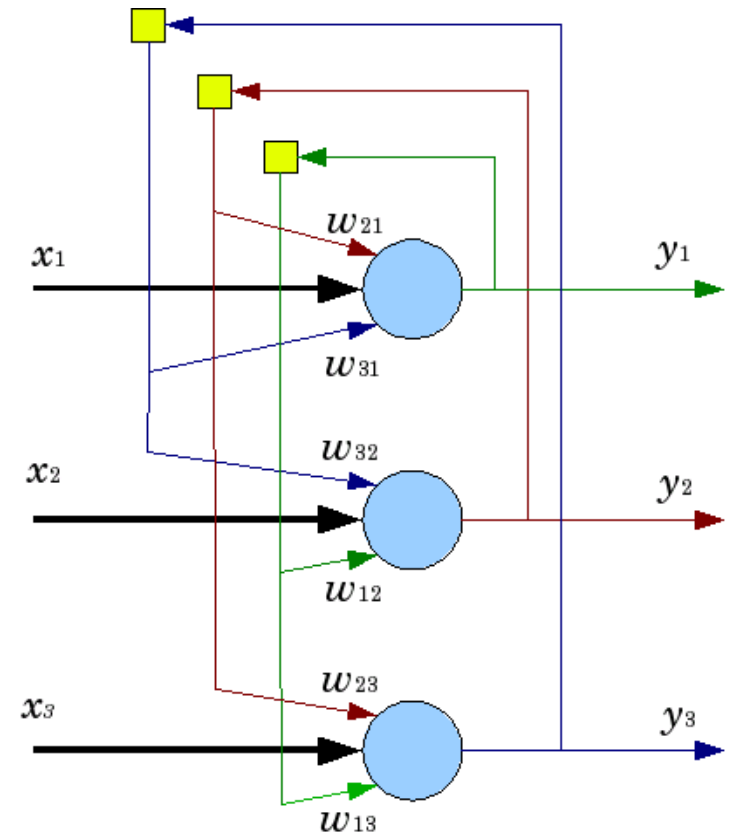


История ML

- **1981. Explanation Based Learning (EBL)**
- Джеральд Дежонг вводит **обучение, основанное на объяснении**, где компьютерный алгоритм анализирует данные и создает общее правило, которому он может следовать и отбрасывать неважные данные.

История ML

- **1982.** Сеть Хопфилда (первая рекуррентная нейросеть, RNN)
- Джон Хопфилд создает сеть Хопфилда, полносвязную нейронную сеть с симметричной матрицей связей.
- В процессе работы динамика таких сетей сходится к одному из положений равновесия. Такая сеть может быть использована как автоассоциативная память, как фильтр, а также для решения некоторых задач оптимизации.
- Сети Хопфилда работают до достижения равновесия, когда следующее состояние сети в точности равно предыдущему.



История ML

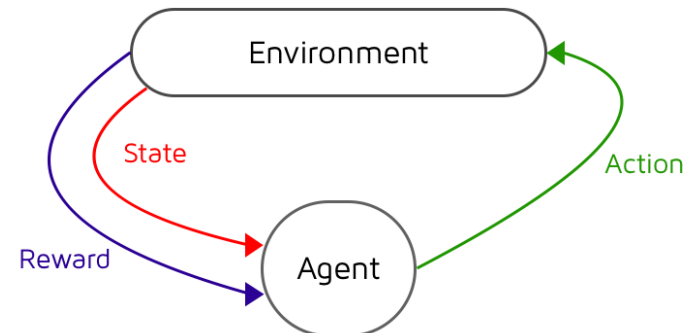
- **1985. Машина Больцмана.**
- Дэвид Х. Экли, Джеффри Хинтон и Терренс Сейновски создают Машину Больцмана, которая является стохастической рекуррентной нейронной сетью. Эта нейронная сеть имеет только входной слой и скрытый слой, но не имеет выходного слоя.
- **1986. NetTalk. Нейросети учатся говорить.**
- Терренс Сейновски создает NetTalk, нейронную сеть, которая учится произносить письменный английский текст.

История ML

- **1986. Реализация Backpropagation.**
- **Джеффри Хинтон, Дэвид Румелхарт и Рональд Уильямс** в статье «Learning Representations by back-propagating errors» показывают успешную реализацию обратного распространения в нейронной сети. Это упростило процесс обучения сложных глубоких нейронных сетей, что было основным препятствием на ранних этапах исследований в этой области.
- Одновременно этот результат получен **С.И. Барцевым** и **В.А. Охониным** (Красноярская группа).
- **Ограниченная машина Больцмана.**
- Пол Смоленски предлагает вариант машины Больцмана, в котором внутри входного и скрытого слоя нет внутрислойной связи. Приобрела популярность только после изобретения Хинтоном быстрых алгоритмов обучения в середине 2000-х годов, особенно широко используется для создания рекомендательных систем.

История ML

- **1989. Q-learning.**
- **Кристофер Уоткинс** развивает **Q-обучение**, что значительно повышает возможности практического применения **обучения с подкреплением**.
- На основе получаемого от среды вознаграждения агент формирует функцию полезности Q , что впоследствии дает ему возможность уже не случайно выбирать стратегию поведения, а учитывать опыт предыдущего взаимодействия со средой.
- Одно из преимуществ Q-обучения — то, что оно в состоянии сравнить ожидаемую полезность доступных действий, не формируя модели окружающей среды.
- Применяется для ситуаций, которые можно представить в виде марковского процесса принятия решений.



История ML

- **1989. Коммерциализация машинного обучения на персональных компьютерах.**
- Axcelis, Inc. выпускает Evolver, первый коммерческий программный пакет, который коммерциализирует использует **генетические алгоритмы**.
- **Генетический алгоритм** — это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе.
- Является разновидностью эволюционных вычислений, с помощью которых решаются оптимизационные задачи с использованием методов естественной эволюции, таких как наследование, мутации, отбор и кроссинговер.
- Относится к **обучению с подкреплением**.

История ML

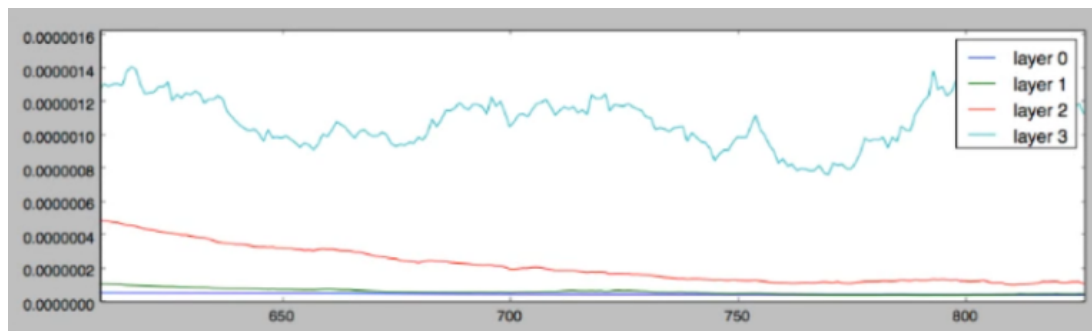
- **1989. CNN использует обратное распространение.**
- **Ян ЛеКун** использует обратное распространение для обучения сверточной нейронной сети для распознавания рукописных цифр. Это прорывной момент, поскольку он закладывает основы современного компьютерного зрения с использованием глубокого обучения.
- **Универсальная теорема аппроксимации.**
- **Джордж Цибенко** публикует самую раннюю версию теоремы об универсальном приближении в своей статье «Приближение суперпозициями сигмоидальной функции». Он доказывает, что прямая нейронная сеть с одним скрытым слоем, содержащим конечное число нейронов, может аппроксимировать любую непрерывную функцию. Это также добавляет доверие к глубокому обучению.

История ML

- **1991. Поиск ассоциативных правил (associations rules learning).**
- В статье «Discovery, analysis and presentation of strong rules». G. Piatetsky-Shapiro. Knowledge Discovery in Databases, AAAI Press, (1991) поставлена задача *поиска ассоциативных правил*.
- «Кто купил x, также купил y». В основе лежит анализ транзакций, каждая из которых представляет собой уникальный набор товаров. При помощи ARL алгоритмов находятся те самые «правила» сочетания товаров внутри одной транзакции, которые потом сортируются

История ML

- **1991. Проблема исчезающего градиента.**
- **Зепп Хохрейтер** определяет проблему исчезающего градиента, которая может сделать обучение глубокой нейронной сети чрезвычайно медленным и практически нецелесообразным. Эта проблема будет продолжать влиять на глубокое обучение еще долгие годы.
- Когда у вас есть глубокая сеть (то есть со многими слоями), при реализации алгоритма обратного распространения градиент стремится к нулю, или, другими словами, исчезает по мере того, как реализуется обратное распространение. В результате коэффициенты не настраиваются, сеть не обучается.



История ML

- **1992. Компьютер играет в нарды.**
- Джеральд Тесавро разрабатывает **TD-Gammon**, компьютерную программу по нардам, в которой используется искусственная нейронная сеть, обученная с использованием **temporal-difference learning** (отсюда и название «TD»). TD-Gammon способен соперничать, но не всегда превосходить способности лучших игроков в нарды.
- **Temporal-difference learning** относятся к обучению с подкреплением.

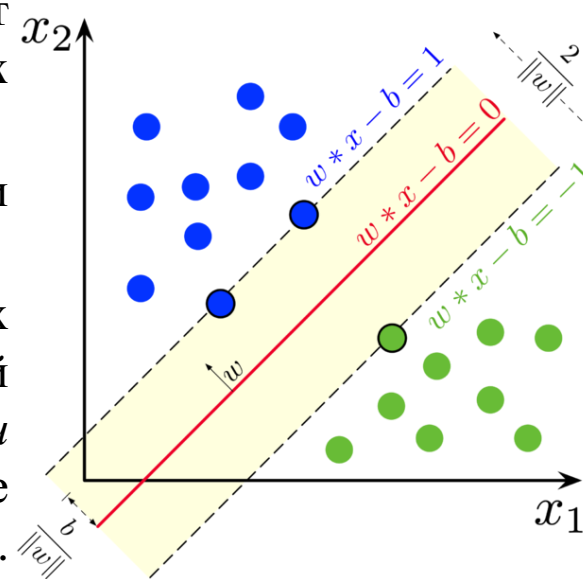


История ML

- **1995. Random decision forest.**
- Тин Кам Хо публикует статью о случайных лесах. Он использует **метод случайных подпространств (random subspace method, attribute bagging or feature bagging)**.
- Это **ансамблевый метод обучения**, который пытается уменьшить корреляцию между оценщиками в ансамбле, обучая их на случайных выборках признаков вместо всего набора признаков. Основная идея заключается в использовании большого ансамбля решающих деревьев, каждое из которых само по себе даёт очень невысокое качество классификации, но за счёт их большого количества результат получается хорошим.
- Хо установил, что леса деревьев, разделенных гиперплоскостями, могут выигрывать в точности, поскольку они растут, не страдая от переобучения, при условии, что леса случайно ограничены и чувствительны только к выбранным характеристикам.

История ML

- **1995. Support vector machine (SVM).**
- Коринна Кортес и Владимир Вапник публикуют свои работы, посвященные **методу опорных векторов**.
- Обучения с учителем, для задач классификации и регрессионного анализа.
- Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск *разделяющей гиперплоскости* с максимальным зазором в этом пространстве между параллельными ей гиперплоскостями. Алгоритм работает в предположении, что чем больше расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.



История ML

- **1997. Long Short-Term Memory.**
- Зепп Хохрейтер и Юрген Шмидхубер публикуют знаковую работу «*Long Short-Term Memory*» (LSTM) (долгая краткосрочная память). Это тип рекуррентной архитектуры нейронных сетей, которая революционизирует глубокое обучение в ближайшие десятилетия.
- LSTM-модуль — это рекуррентный модуль сети, способный запоминать значения как на коротких, так и на длинных промежутках времени.
- Ключом к данной возможности является то, что LSTM-модуль не использует функцию активации внутри своих рекуррентных компонентов. Таким образом, хранимое значение не размывается во времени, и градиент не исчезает при использовании метода обратного распространения ошибки во времени при обучении сети.

История ML

- **1997. Deep Blue от IBM побеждает чемпиона мира по шахматам.**
- **Deep Blue** — шахматный суперкомпьютер, разработанный компанией IBM, который 11 мая 1997 года выиграл матч из 6 партий у чемпиона мира по шахматам **Гарри Каспарова (3.5:2.5)**.
- В 1996 году компьютер проиграл 2:4.
- Гарри Каспаров обвинил организаторов матча в нечестной игре, заявив, что компьютер получал помощь от человека. Позже, в интервью 2016 года, он заявил, что пересмотрел свои взгляды по этому поводу.



История ML

- **1998. MNIST.**
- Команда во главе с Янном ЛеКуном выпускает базу данных MNIST, набор данных, состоящий из рукописных цифр сотрудников Американского бюро переписи и американских школьников.
- С тех пор база данных MNIST стала эталоном для оценки распознавания рукописного текста.



История ML

- **2001. Random decision forest.**
- Расширение алгоритма случайных лесов было разработано Лео Брейманом и Аделью Катлер, которые зарегистрировали «Random forest» в качестве товарного знака (по состоянию на 2019 год, принадлежащий Minitab, Inc.).
- *Breiman L (2001). "Random Forests". Machine Learning. 45 (1): 5–32.* В этой статье описывается метод построения леса некоррелированных деревьев с использованием процедуры, подобной CART (Classification and Regression Tree), в сочетании с рандомизированной оптимизацией узлов и формированием ансамблей.
- Алгоритм сочетает в себе две основные идеи: метод **бэггинга** Бреймана, и метод случайных подпространств, предложенный Хо. Алгоритм применяется для задач классификации, регрессии и кластеризации.

Современность ML

- **2002. Torch Machine Learning Library**
- Первый выпуск Torch, библиотеки программного обеспечения для машинного обучения.
- **2006. Netflix Prize**
- Объявлен конкурс компании Netflix. Цель конкурса состояла в том, чтобы использовать машинное обучение для улучшения точности рекомендаций программного обеспечения Netflix в прогнозировании рейтинга пользователя для фильма, учитывая его оценки для предыдущих фильмов.
- Главный приз составлял \$1,000,000. Для его получения необходимо было улучшить алгоритм Netflix на 10 %.
- Приз был выдан команде BellKor's Pragmatic Chaos 21 сентября 2009 года. Команда The Ensemble получила такой же результат, но прислала решение на 20 минут позже.

Современность ML

- **2006. Deep Belief Network.**
- Джеффри Хинтон, Руслан Салахутдинов, Осиндеро и Тех публикуют статью “*A fast learning algorithm for deep belief nets*”.
- **2008. Революция GPU.**
- Группа Эндрю Нг в Стэнфорде начинает пропагандировать использование графических процессоров для обучения глубоких нейросетей, чтобы в разы ускорить время обучения. Это может принести практическую пользу в области глубокого обучения для эффективного обучения огромному объему данных.

Современность ML

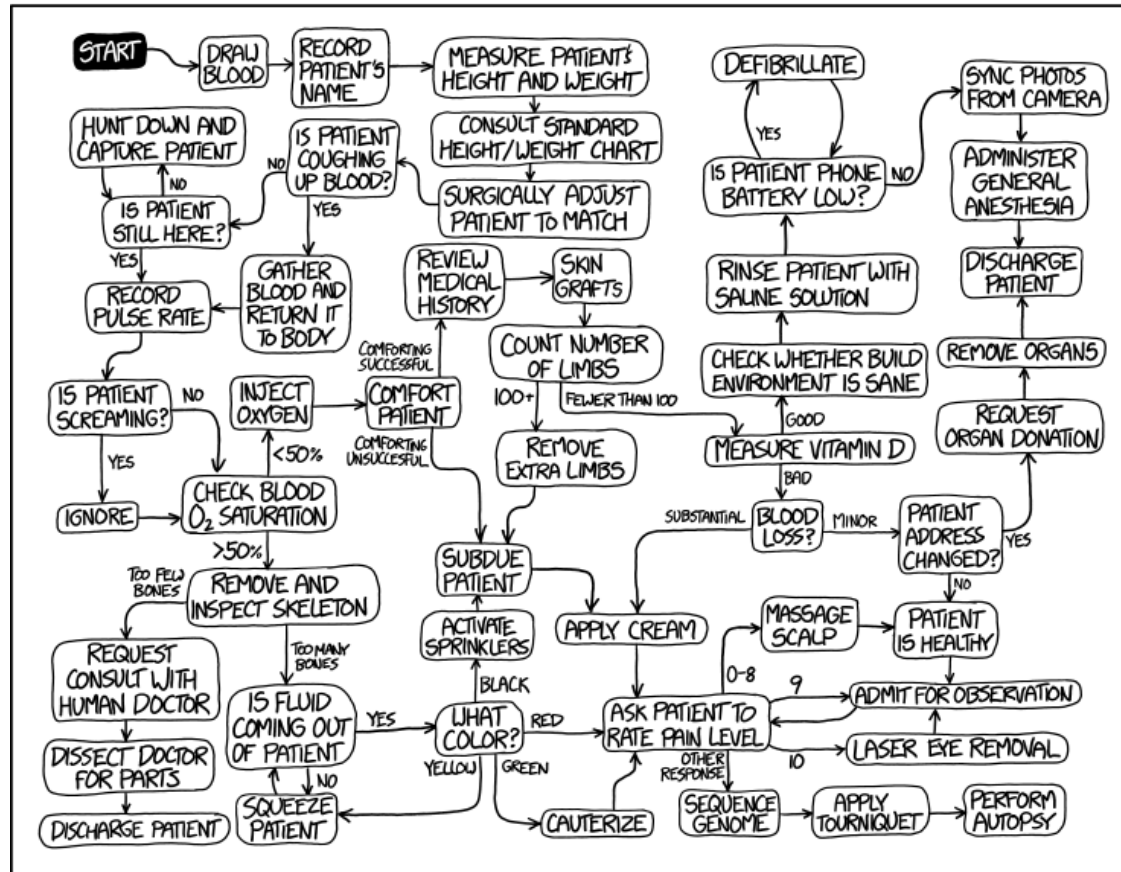
- **2009. ImageNet.**
- ImageNet – это большая база данных изображений, созданная Фэй-Фей Ли из Стэнфордского университета, которая поняла, что лучшие алгоритмы машинного обучения не будут работать хорошо, если данные не отражают реальный мир. Для многих ImageNet стал катализатором бума искусственного интеллекта 21-го века.
- На август 2017 года в ImageNet 14 197 122 изображения, разбитых на 21 841 категорию.
- С 2010 года ведётся проект ILSVRC (*ImageNet Large Scale Visual Recognition Challenge* — Кампания по широкомасштабному распознаванию образов в ImageNet), в рамках которого различные программные продукты ежегодно соревнуются в классификации и распознавании объектов и сцен в базе данных ImageNet.

Современность ML

- **2010. Kaggle.**
- Запущен веб-сайт Kaggle, который служит платформой для соревнований по машинному обучению
- **2011. Сражение с исчезающим градиентом.**
- Йошуа Бенжио, Антуан Бордес, Ксавье Глорот в своей статье «*Deep Sparse Rectifier Neural Networks*» показывают, что функция активации ReLU позволяет избежать исчезающей проблемы градиента. Это означает, что теперь, кроме GPU, у сообщества глубокого обучения есть еще один инструмент, позволяющий избежать проблем с длительным временем обучения глубокой нейронной сети.
- **2011. «Jeopardy!»**
- Используя комбинацию машинного обучения, обработки естественного языка и методов поиска информации, компьютер Watson (IBM) побеждает двух чемпионов в «Jeopardy!»

Современность ML

A GUIDE TO THE MEDICAL DIAGNOSTIC AND TREATMENT ALGORITHM USED BY IBM'S WATSON COMPUTER SYSTEM



Современность ML

- **2012. Котики! Бум глубокого обучения.**
- Команда Google Brain, возглавляемая Эндрю Нг и Джеффом Дином, создает нейронную сеть, которая учится распознавать кошек, просматривая немаркированные изображения, взятые из кадров видео YouTube.
- AlexNet, сверточная нейронная сеть, разработанная Алексом Крижевским, побеждает в конкурсе изображений Imagenet с точностью до 84%. Это огромный скачок от 75% точности, достигнутой более ранними моделями. Эта победа вызывает новый бум глубокого обучения во всем мире.
- Конкурс ImageNet состоялся в октябре 2012 года и был посвящен классификации объектов на фотографиях. В конкурсе требовалось распознавание образов в 1000 категорий.

Современность ML

Цветок из обучающей выборки



Система обнаружила цветы



Современность ML

**Система распознала белого
медведя**



**Система распознала медведя
гризли**



Современность ML

**Банановый слизняк, ошибочно
распознан как змея**



Ослик, распознан как собака



Современность ML

- **2014.GAN**
- **Generative Adversarial Neural Network**, также известная как GAN (генеративно-сопоставительная сеть), создана Яном Гудфеллоу. GAN открывают новые возможности применения глубокого обучения в моде, искусстве, науке благодаря своей способности синтезировать данные, похожие на реальные.
- Сеть построена на комбинации из двух нейронных сетей, одна из которых (сеть G) генерирует образцы, а другая (сеть D) старается отличить правильные («подлинные») образцы от неправильных. Так как сети G и D имеют противоположные цели — создать образцы и отбраковать образцы — между ними возникает Антагонистическая игра.
- GAN используются для получения фотореалистичных изображений, например для элементов промышленного дизайна, дизайна интерьера, одежды, сумок, портфелей, сцен компьютерных игр и т. д. Также они помогают воссоздать трёхмерную модель объекта из фрагментов и улучшить изображения, полученные из астрономических наблюдений.

Современность ML



Современность ML



Современность ML

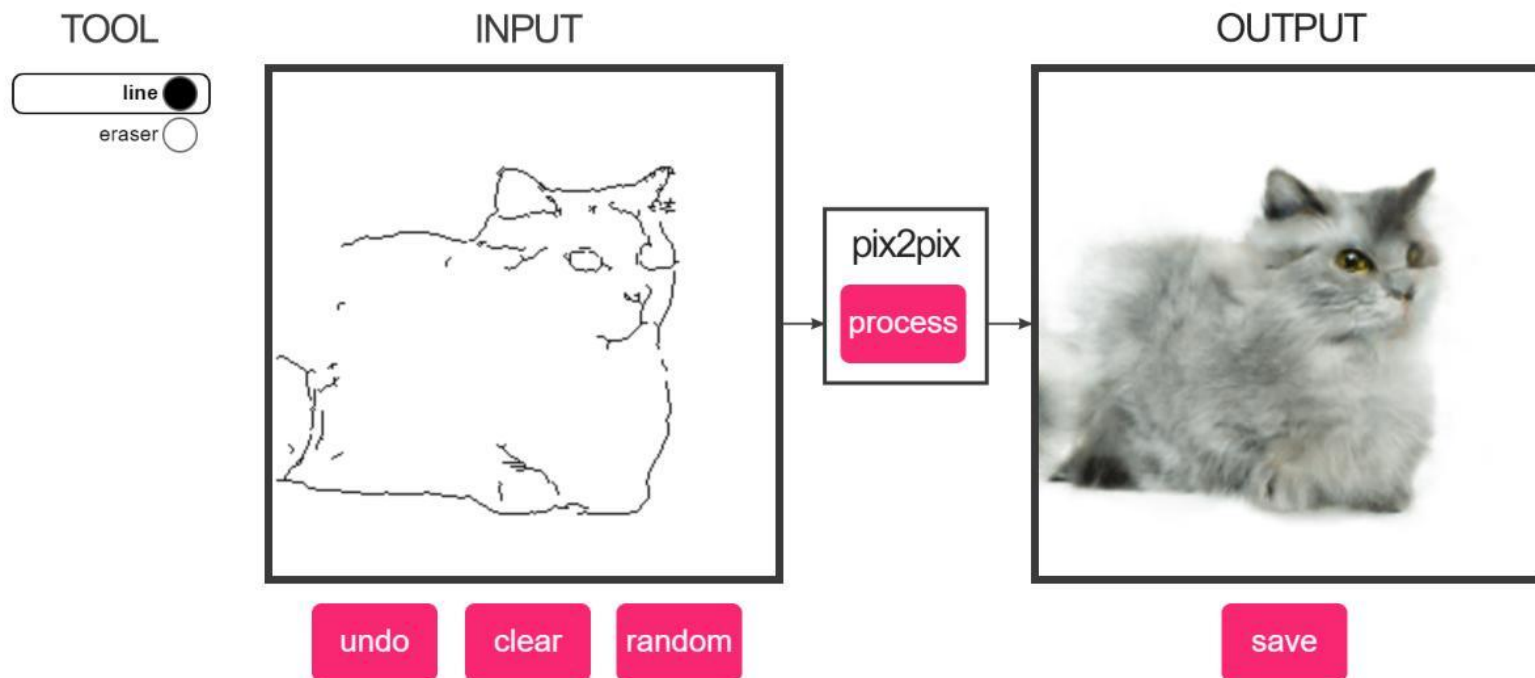
Source Image				
CycleGAN				
UNIT				

Современность ML



Современность ML

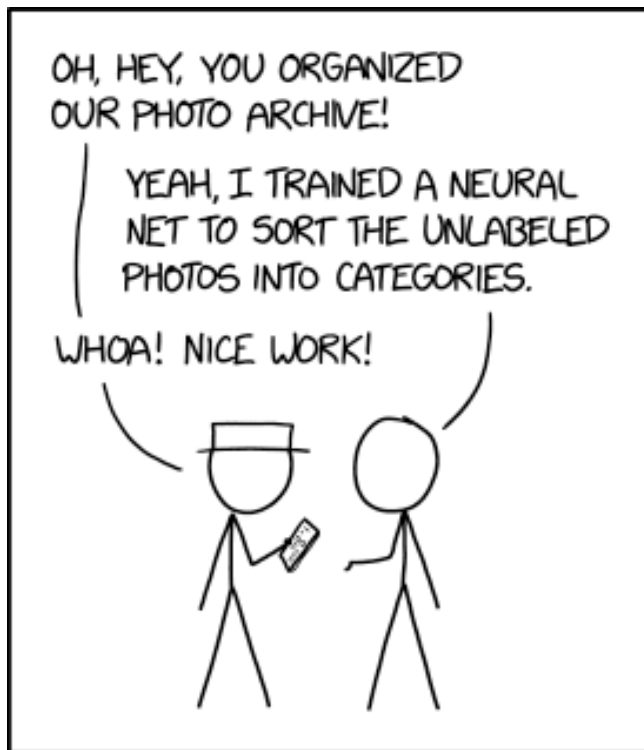
edges2cats



Современность ML

- **2014. DeepFace.**
- Исследователи Facebook публикуют свои работы по DeepFace, системе, которая использует нейронные сети для идентификации лиц. Результатом является улучшение более чем на 27% по сравнению с предыдущими системами (до 97,35%).
- **2014. Sibyl.**
- Исследователи из Google подробно рассказывают о своей работе над Sibyl, собственной платформой для массового параллельного машинного обучения, используемой Google для прогнозирования поведения пользователей и предоставления рекомендаций.
- **2014 г. Чатбот Женя Густман проходит тест Тьюринга.**
- Он убеждает 33% экспертов в том, что является 13-летним мальчиком из Одессы. Эксперты пять минут переписываются одновременно с живым человеком и роботом, находясь в разных комнатах и не видя друг друга. Потом каждый эксперт говорит, кто был человеком, а кто — программой

Современность ML



ENGINEERING TIP:
WHEN YOU DO A TASK BY HAND,
YOU CAN TECHNICALLY SAY YOU
TRAINED A NEURAL NET TO DO IT.



Современность ML

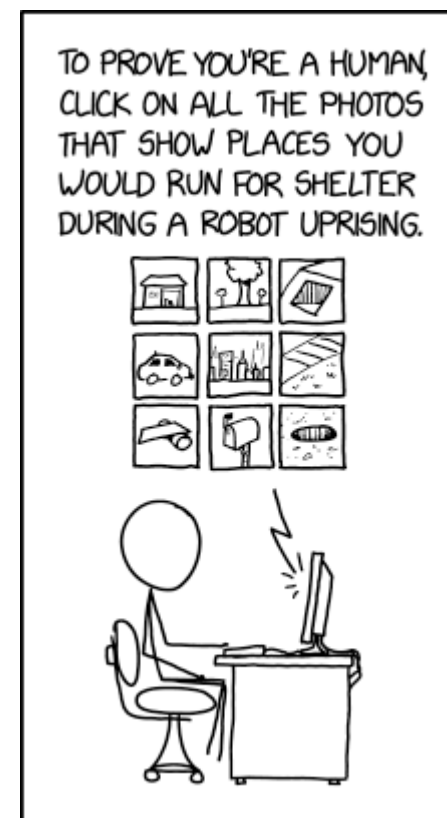
- **2015 г. Машины и люди объединяются для борьбы с мошенничеством в интернете.**
- Когда PayPal собрался бороться с мошенничеством и отмыванием денег через сервис, он применил смешанный подход. Детективы задают параметры преступного поведения, а машина с помощью этих параметров выявляет мошенников на сайте PayPal.
- **2016 г. Компьютер научился читать по губам.**
- HAL 9000, вымышленный компьютер из «Космической одиссеи 2001 года», умел читать по губам. Команда из Оксфорда сумела создать нейросеть LipNet, которая распознает слова по движениям губ с точностью 93.4%. Таким образом она превратила научную фантастику в реальность.

Современность ML

- **2016. DeepFake.**
- **Deepfake** (дипфейк) — конкатенация слов «глубокое обучение» (Deep learning) и «подделка» (Fake), методика синтеза изображения, основанная на искусственном интеллекте. Она используется для соединения и наложения существующих изображений и видео на исходные изображения или видеоролики.
- Технология deepfake использовалась в музейном пространстве, например, в Музее Сальвадора Дали во Флориде в честь 115-летия известного художника была организована специальная выставка Dalí Lives («Дали жив»). Они использовали прототип художника, который общался с посетителями музея, рассказывал им истории о своих картинах и жизни, а также делал селфи и присылал фотографии на почту посетителей.
- Также используется создания фальшивых порнографических видео. Deepfake-порнография появилась в Интернете в 2017 году, в частности на Reddit и была запрещена сайтами, включая Reddit, Twitter и Pornhub.

Современность ML

- **2016 г. Обработка естественного языка дает жизнь цифровому персональному консультанту.**
- North Face стала первым продавцом, которая использовала IBM Watson для обработки естественного языка в мобильном приложении. Цифровой персональный консультант общается с клиентами и помогает им найти нужную вещь.
- **2017 г. Машина научилась распознавать онлайн-троллинг.** Компания Jigsaw создала систему, которая научилась определять троллинг, прочитав миллион комментариев на сайте. Алгоритмы системы могут пригодиться владельцам сайтов, у которых не хватает ресурсов на модерацию.



Современность ML

REVIEW OF CONVERSATION

HIGH MATCH

MEN'S THERMOBALL™ SP TRICLIMATE® PARK
\$349.00

HIGH MATCH

MEN'S GATEKEEPER 2.7 JACKET
\$299.00

HIGH MATCH

MEN'S POINT IT DOWN HYBRID JACKET
\$380.00

HIGH MATCH

MEN'S VORTEX TRICLIMATE® JACKET
\$280.00

HIGH MATCH

MEN'S MARSELL TRICLIMATE® JACKET
\$280.00

WHAT COLOR JACKET ARE YOU LOOKING FOR?

Современность ML

- **2017. AlphaGo победила человечество в го**
- Программа AlphaGo победила в матче против Кэ Цзе, сильнейшего в мире игрока в го. AlphaGo выиграла вторую игру из запланированных трех.
- В 2015 году программа AlphaGo обыграла чемпиона Европы, а в марте 2016 года продемонстрировала высокий уровень игры, победив Ли Седоля, одного из сильнейших игроков в го в мире.
- В **2015** году математик **Джон Тромп** завершил подсчет числа возможных не противоречащих правилам комбинаций камней в го при игре на стандартном гобане размером 19×19 . Получившееся число оказалось больше, чем количество атомов в наблюдаемой Вселенной

Современность ML



Современность ML

- **2019. Премия Тьюринга.**
- Йошуа Бенжио, Джеффри Хинтон и Янн ЛеКун получают премию Тьюринга 2018 за их огромный вклад в достижения в области глубокого обучения и искусственного интеллекта. Это определяющий момент для тех, кто неустанно работал над нейронными сетями, когда все сообщество по машинному обучению отошло от него в 1970-х годах.

