Тема 1.Введение. Основные понятия машинного обучения. Применение машинного обучения в искусственном интеллекте

Что же такое машинное обучение?

Машинное обучение считается ветвью искусственного интеллекта, основная идея которого заключается в том, чтобы компьютер не просто использовал заранее написанный алгоритм, а сам обучился решению поставленной задачи.

Понятие машинного обучения

Машинное обучение (machine learning) — подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин.

Виды машинного обучения

Обучение по прецедентам (индуктивное обучение) основано на выявлении общих закономерностей по частным эмпирическим данным.

Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Компания IBM внесла немалый вклад в историюКомпания IBM внесла немалый вклад в историю машинного обучения. Так, ввод в обиход термина «машинное обучение» приписывают одному из сотрудников компании, Артуру Самюэлю в его исследовании игры в шашки. В 1962 году самопровозглашенный мастер по шашкам Роберт Нили сыграл партию с компьютером IBM 7094 и проиграл. По сравнению с современными возможностями это достижение кажется сущим пустяком, но оно считается важной вехой в области искусственного интеллекта. В следующие пару десятилетий технологии в области хранения данных и вычислительные мощности достигнут такого уровня, что будут созданы революционные в то время (но привычные и любимые сегодня) продукты, например система рекомендаций Netflix или беспилотные автомобили.

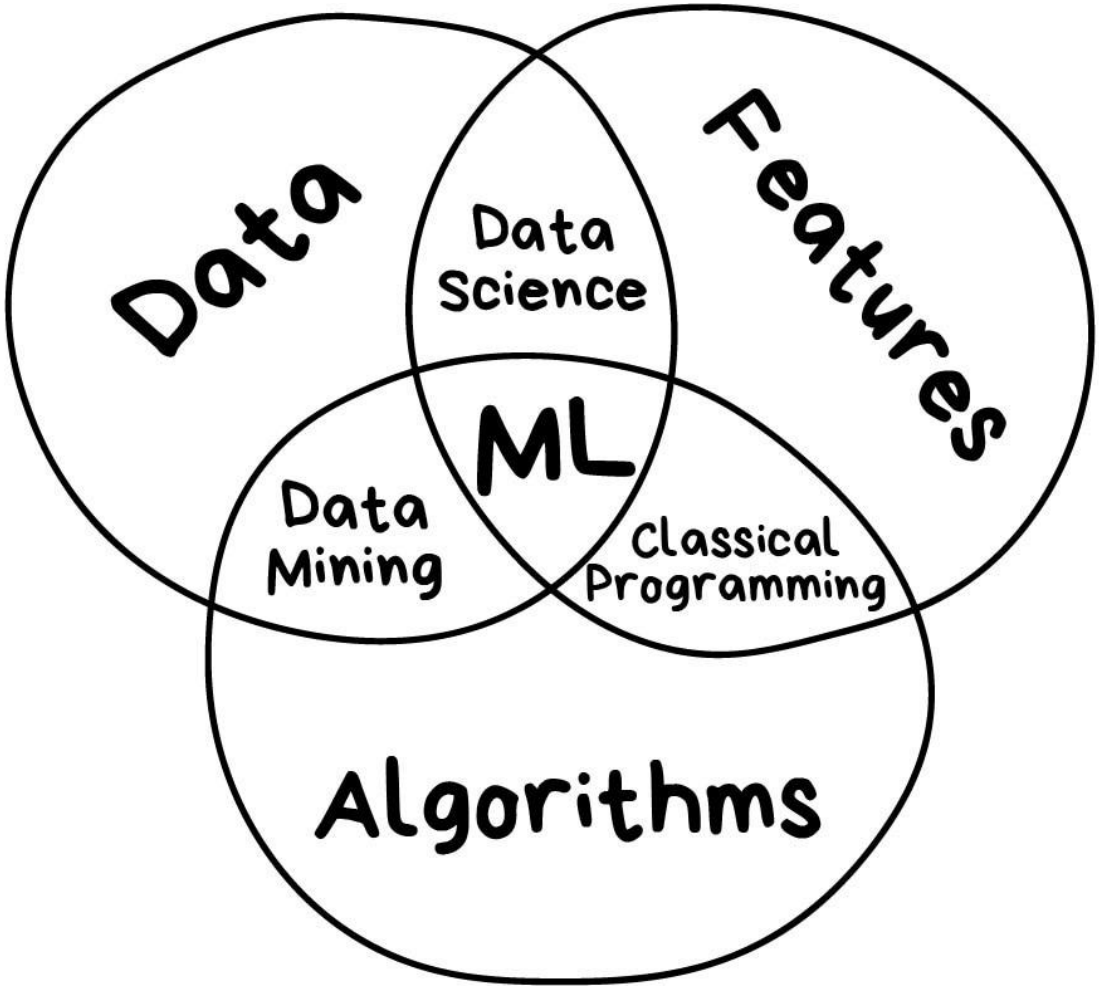
История развития машинного обучения

- ***1997*** Компьютер Deep Blue обыграл чемпиона мира по шахматам Гарри Каспарова.
- ***2006*** Джеффри Хинтон (Geoffrey Hinton), ученый в области искусственных нейросетей, ввел в обиход термин «Глубинное обучение» (Deep learning).
- ***2011*** Эндрю Нг (Andrew Ng) и Джефф Дин (Jeff Dean) основали Google Brain.
- ***2011*** Суперкомпьютер IBM Watson, оснащенный системой искусственного интеллекта, одержал победу в телевикторине Jeopardy!. Его соперниками были маститые игроки Брэд Раттер (Bred Ratter) и Кен Дженнингс (Ken Jennings).
- ***2012*** В Google X Lab разработали алгоритм, позволяющий идентифицировать видиоролики, содержащие котов.

Так как люди часто путают глубокое обучение и машинное обучение, давайте остановимся на отличительных особенностях каждого из этих понятий. Машинное обучение, глубокое обучение и нейронные сети — все это подразделы искусственного интеллекта. Но при этом глубокое обучение является подвидом машинного обучения, а нейронные сети, в свою очередь, — подвидом глубокого обучения.

Разница между глубоким и машинным обучением заключается в способе обучения алгоритмов. В глубоком обучении большая часть процесса извлечения признаков автоматизирована, что практически исключает необходимость контроля со стороны человека и позволяет использовать большие наборы данных. Лекс Фридман называет глубокое обучение «масштабируемым машинным обучением». Эффективность классического, «неглубокого» машинного обучения в большей степени зависит от контроля со стороны человека. Набор признаков для понимания разницы между входными данными определяется специалистом-человеком. Обычно для машинного обучения требуются более структурированные данные.

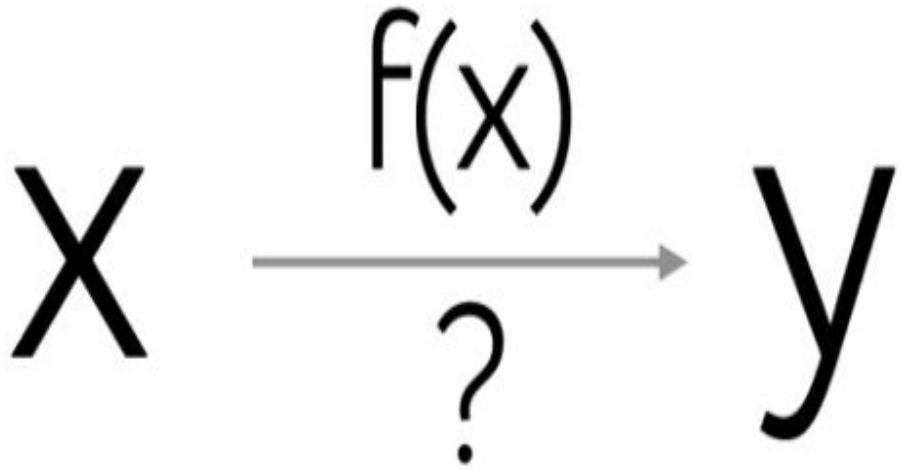
Три составляющие обучения



Карта мира машинного

Если у нас есть какие то данные о наших *X* *и соответствующие к ним данные* *Y* *, мы можем попытаться, зависимость между ними приблизить. Нас совершенно устроит, что это приближение не будет идеальным*

Пример: отзывы в интернете



Текст отзыва

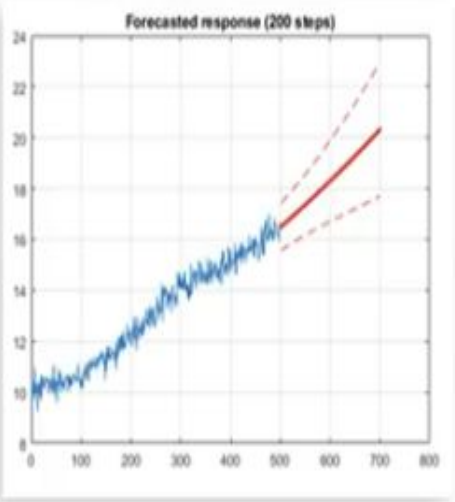
Окрас отзыва

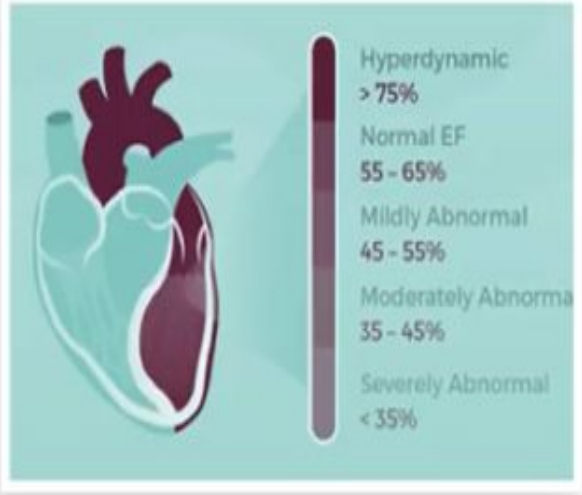
«Достаточно неплохое видео, кажется, я даже что-то понял»

«То, что вы делаете – полный отстой!»

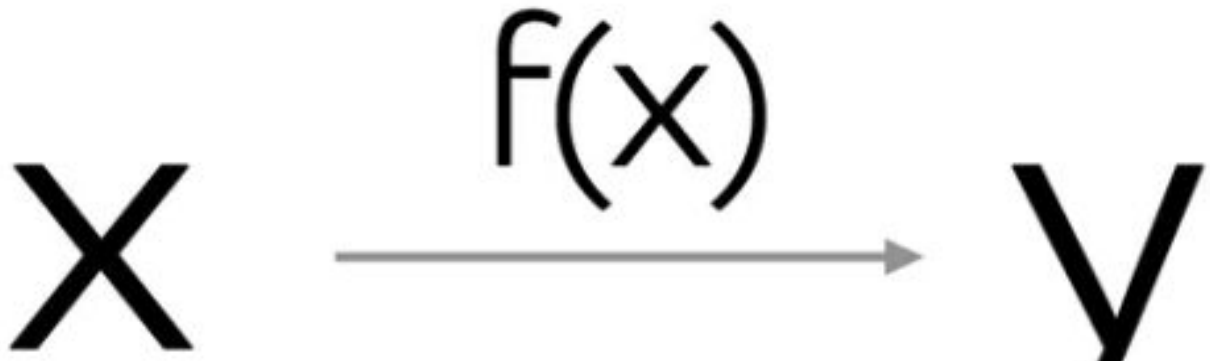
Больше сложных зависимостей!

$$x \xrightarrow{f(x)} y$$





Суть машинного обучения



$$y \approx f(x)$$

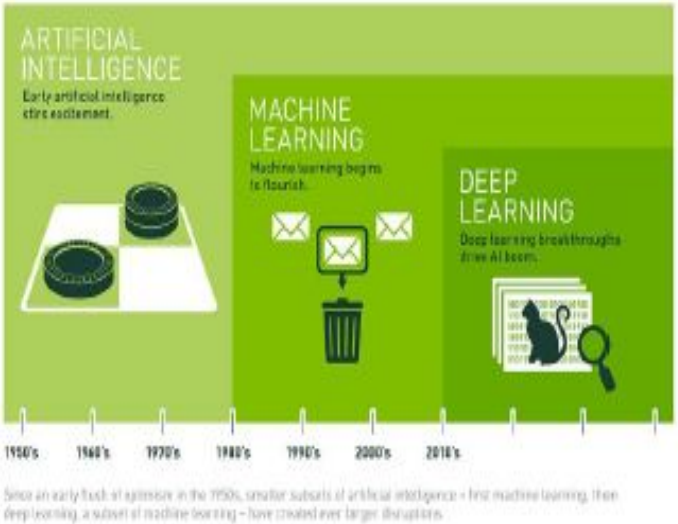
Суть машинного обучения

Мы можем приближать сложные зависимости и даже модели не имея не малейшего понятия как они устроены, в отличие физики и механики

$$y \approx F(x)$$

Машинное обучение (Machine Learning, ML)

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



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

Как работает машинное обучение

Согласно UC Berkeley, система обучения алгоритма машинного обучения состоит из трех основных частей.

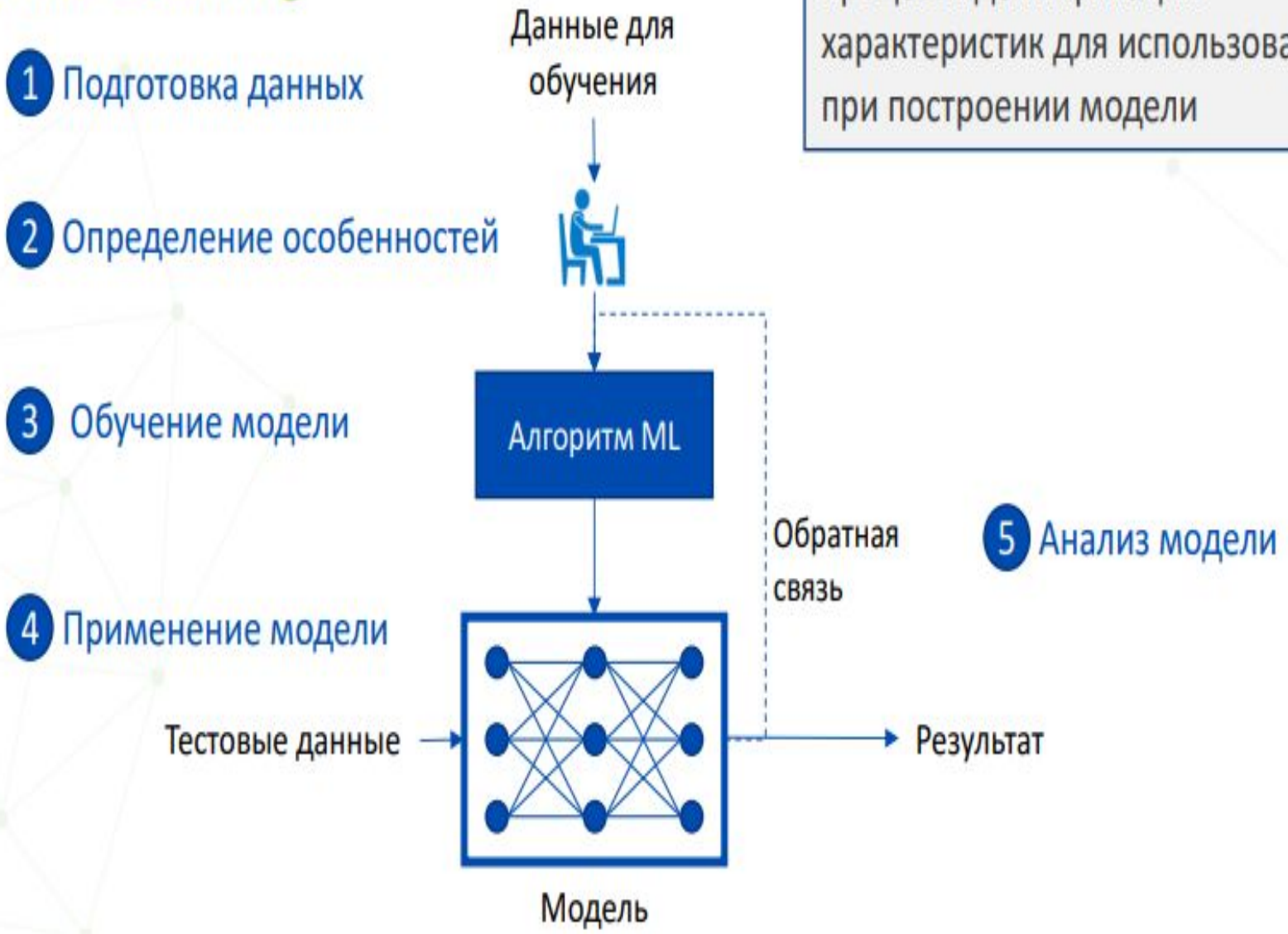
Процесс принятия решений. Как правило, алгоритмы машинного обучения используются для создания прогнозов или классификации данных. Взяв за основу некоторые входные данные, которые могут быть размечены или нет, алгоритм выдаст оценку в отношении наличия закономерности в данных.

Функция ошибок. Функция ошибок служит для оценки прогноза по модели. При наличии известных примеров функция ошибок способна сравнить их, чтобы оценить точность модели.

Процесс оптимизации модели. Если можно еще точнее сопоставить модель с точками данных в учебном наборе, то веса корректируются с целью уменьшения расхождения между известным примером и оценкой модели. Алгоритм будет повторять это вычисление и оптимизировать процесс, самостоятельно обновляя веса до тех пор, пока не будет достигнут порог точности.

Машинное обучение
Machine Learning

Определение особенностей, - это процесс идентификации характеристик для использования при построении модели

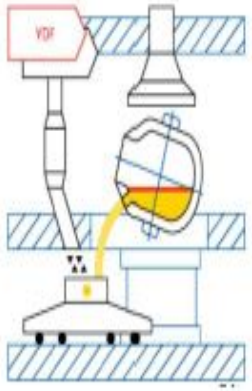


Примеры задач машинного обучения

- **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

- **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 с данными сложной структуры

- **Статистический машинный перевод:**
  объект – предложение на естественном языке
  ответ – его перевод на другой язык

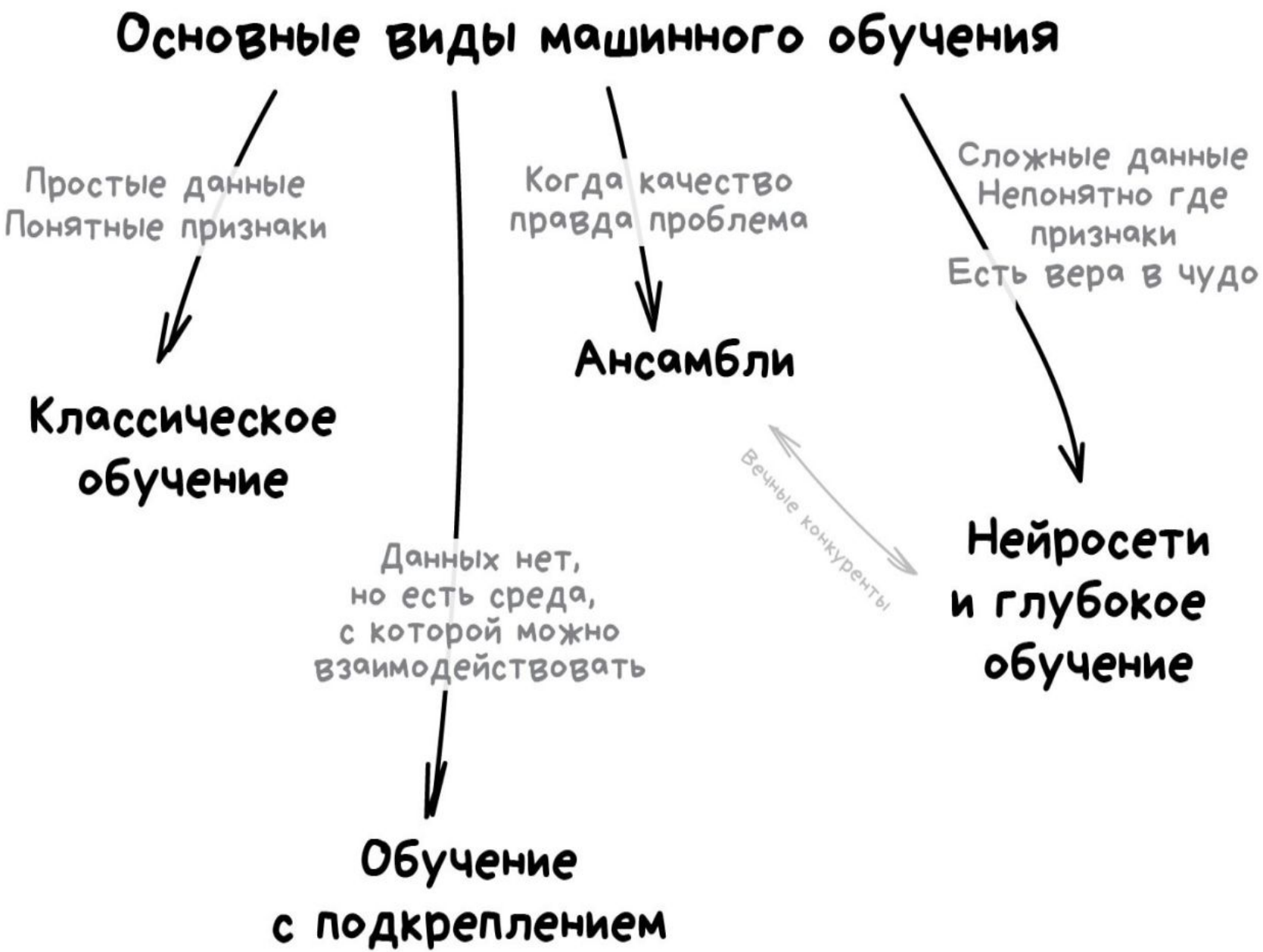
*Прогресс в этих областях связан с «**большими данными**» (англ. «Big Data»)*

- **Перевод речи в текст:**
  объект – аудиозапись речи человека
  ответ – текстовая запись речи

...очень важное уточнение:

*с **аккуратными** большими данными*

- **Компьютерное зрение:**
  объект – динамика сцены в видеопоследовательности
  ответ – решение (объехать, остановиться, игнорировать)


Основные виды машинного обучения
Простые данные
Понятные признаки
Классическое обучение
Данных нет,
но есть среда,
с которой можно
взаимодействовать
Обучение
с подкреплением
Когда качество
правда проблема
Ансамбли
Сложные данные
Непонятно где
признаки
Есть вера в чудо
Нейросети
и глубокое
обучение
Вечные конкуренты

Классическое Обучение

Данные заранее категоризированы или численные

Данные никак не размечены

С учителем

Предсказать категорию

Предсказать значение

Классификация

«Разложи носки по цвету»

Регрессия

«Разложи галстуки по длине»

Без учителя

Разделить по схожести

Найти зависимости

Выявить последовательности

Кластеризация

«Разложи похожие вещи по кучкам»

Ассоциация

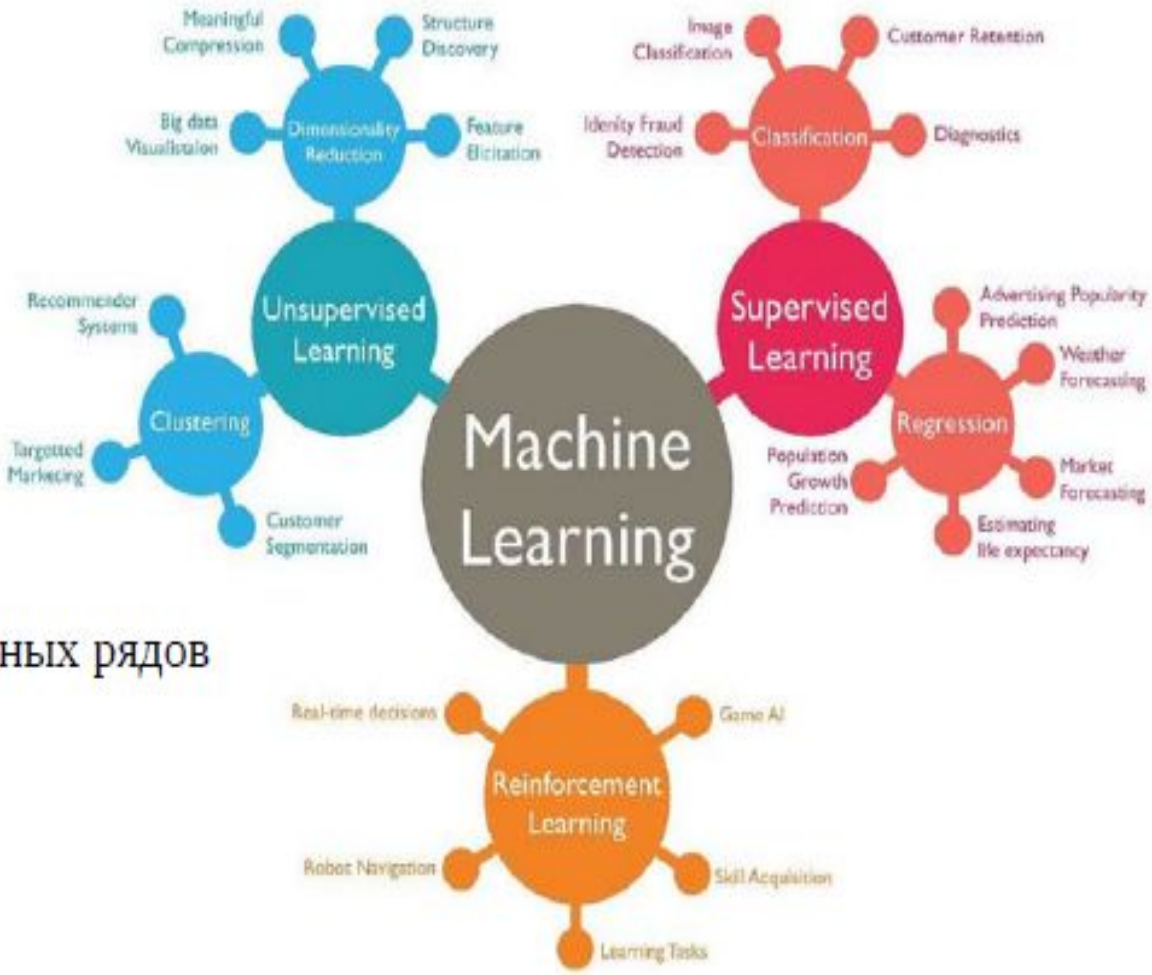
«Найди какие вещи я часто ношу вместе»

Уменьшение Размерности (обобщение)

«Собери из вещей лучшие наряды»

Основные направления ML

- Обучение без учителя
- Обучение с учителем
  - классификация
  - регрессия
  - генеративные модели
  - восстановление временных рядов
  - ...
- Обучение с подкреплением
- ...



Способы машинного обучения

- Обучение с учителем
  - Англ: Supervised learning
  - Позволяет приложениям обучаться без специфического программирования
  - Обучение и прогнозирование на основании известных данных
  - В основе статистика и вероятность
- Сценарии
  - Предсказательная аналитика

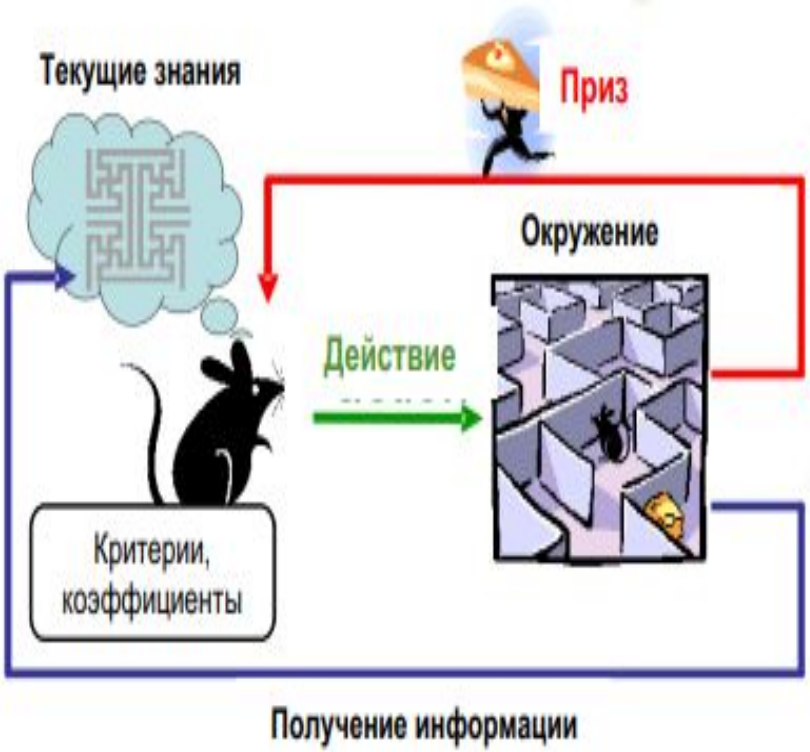
- Обучение без учителя
  - Англ: Unsupervised learning
  - Возможно использование нескольких уровней данных
  - Обработка исходных данных и выдача модифицированного результата на следующий уровень
- Сценарии:
  - Распознавание образов

Способы

- **Обучение с учителем** - для каждого прецедента задаётся пара «ситуация, требуемое решение»:
  - Метод коррекции ошибки
  - Метод обратного распространения ошибки
- **Обучение без учителя** - для каждого прецедента задаётся только «ситуация», требуется сгруппировать объекты в кластеры, используя данные о попарном сходстве объектов, и/или понизить размерность данных:
  - Метод ближайших соседей
- **Обучение с подкреплением** - для каждого прецедента имеется пара «ситуация, принятое решение»:
  - Генетические алгоритмы
  - Альфа-система подкрепления
  - Гамма-система подкрепления

Обучение с подкреплением
Reinforcement Learning

- Частный случай обучения с учителем
- Учитель - среда
- Наличие обратной связи
- Алгоритм проб и ошибок

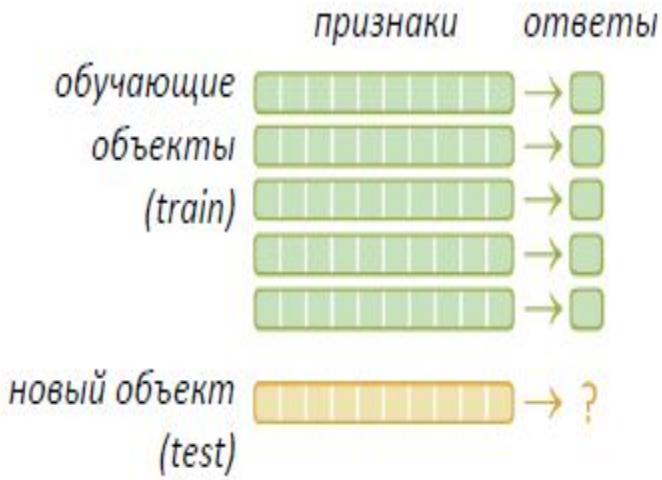


Задача машинного обучения с учителем

Этап №1 – обучение с учителем

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

Если нет данных, то нет и машинного обучения



Этап №2 – применение

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

Обучение с учителем

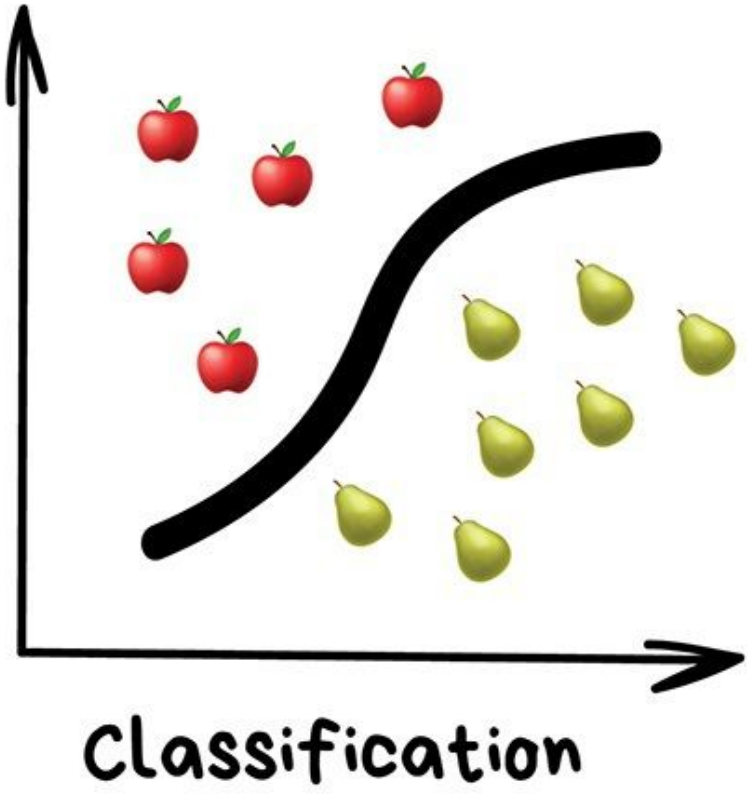
Испытуемая система принудительно обучается с помощью примеров «стимул-реакция». Между входами и эталонными выходами (стимул-реакция) может существовать некоторая зависимость, но она не известна. Известна только конечная совокупность прецедентов — пар «стимул-реакция», называемая обучающей выборкой. На основе этих данных требуется восстановить зависимость (построить модель отношений стимул-реакция, пригодных для прогнозирования), то есть построить алгоритм, способный для любого объекта выдать достаточно точный ответ. Для измерения точности ответов, так же как и в обучении на примерах может вводится функционал качества.

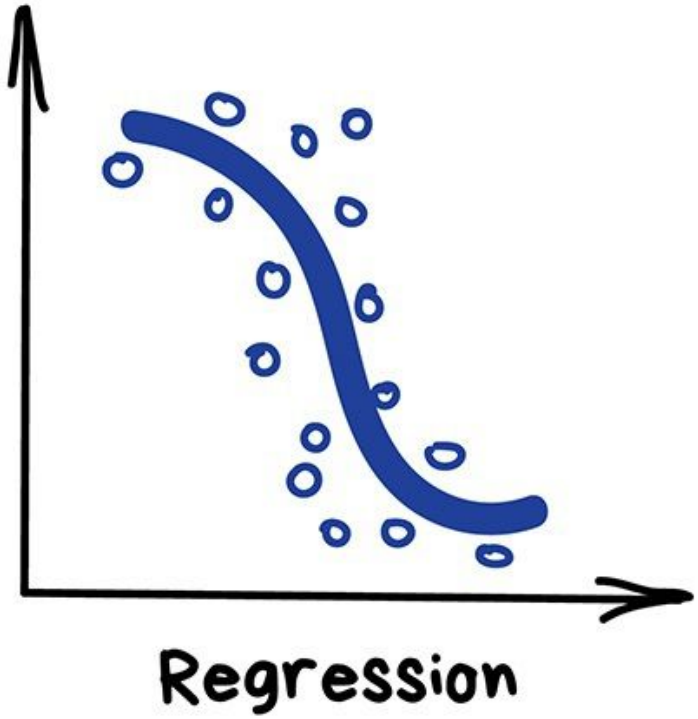
Вырожденные виды «учителей»

Система подкрепления с управлением по реакции (R — управляемая система) — характеризуется, тем что информационный канал от внешней среды к системе подкрепления не функционирует. Данная система несмотря на наличие системы управления относится к спонтанному обучению, так как испытуемая система обучается автономно, под действием лишь своих выходных сигналов независимо от их «правильности». При таком методе обучения для управления изменением состояния памяти не требуется никакой внешней информации;

Система подкрепления с управлением по стимулам (S — управляемая система) — характеризуется, тем что информационный канал от испытываемой системы к системе подкрепления не функционирует. Несмотря на не функционирующий канал от выходов испытываемой системы относится к обучению с учителем, так как в этом случае система подкрепления (учитель) заставляет испытываемую систему вырабатывать реакции согласно определенному правилу, хотя и не принимается во внимание наличие истиных реакций испытываемой системы.

Задачи классификации и регрессии

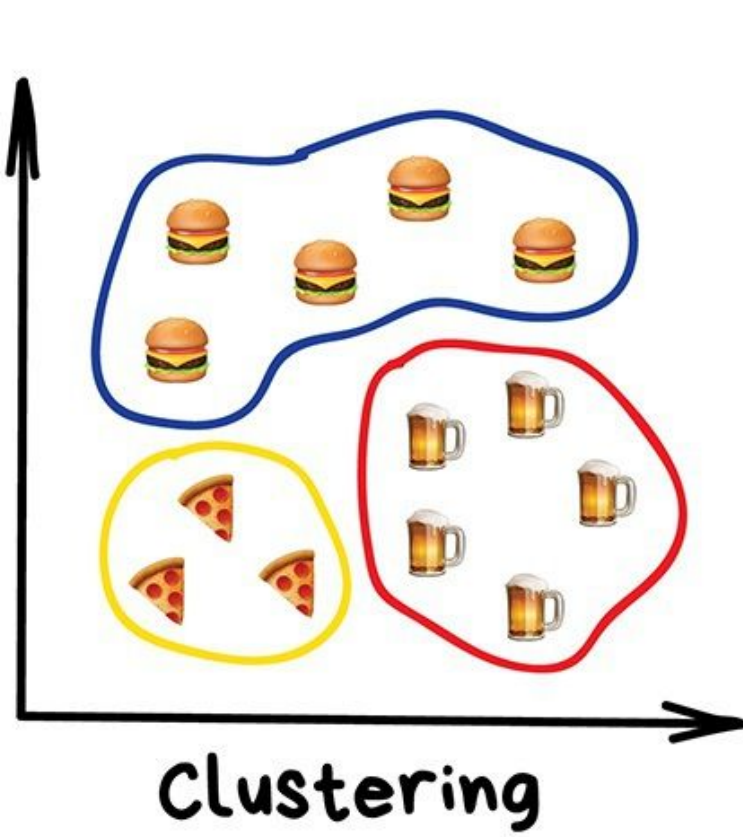


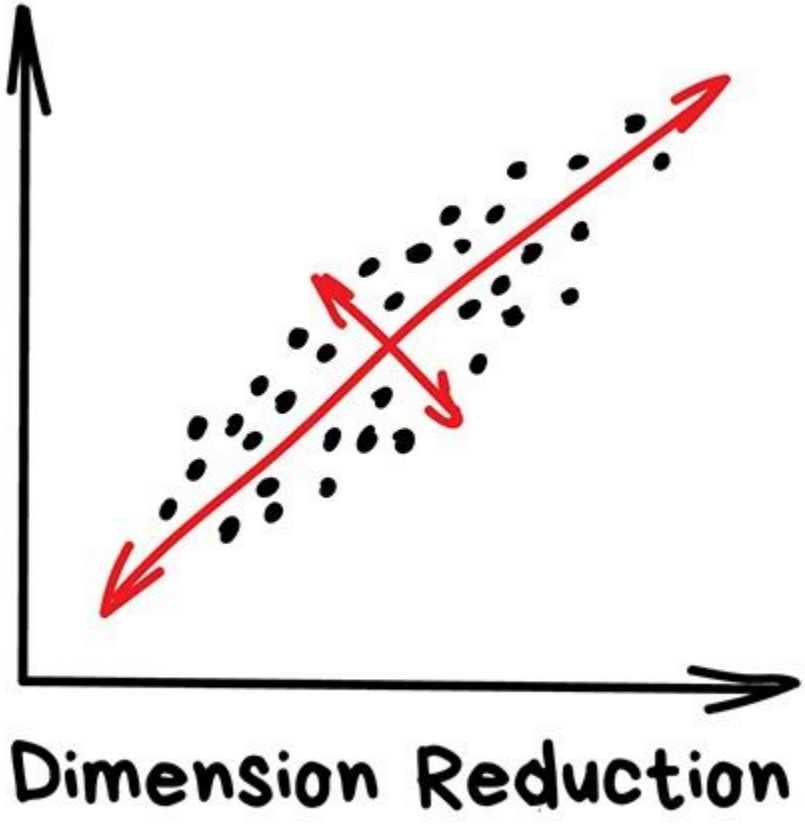


Обучение без учителя

- Испытуемая система спонтанно обучается выполнять поставленную задачу, без вмешательства со стороны экспериментатора.
- Как правило, это пригодно только для задач, в которых известны описания множества объектов (обучающей выборки), и требуется обнаружить внутренние взаимосвязи, зависимости, закономерности, существующие между объектами.

Кластеризация и уменьшение размерности (абстракция)





Обучение с подкреплением

- Испытуемая система (агент) обучается, взаимодействуя с некоторой средой. Откликом среды (а не специальной системы управления подкреплением, как это происходит в обучении с учителем) на принятые решения являются сигналы подкрепления, поэтому такое обучение является частным случаем обучения с учителем, но учителем является среда или ее модель.
- Также нужно иметь в виду, что некоторые правила подкрепления базируются на неявных учителях, например, в случае ИНС, на одновременной активности формальных нейронов, из-за чего их можно отнести к обучению без учителя.

Альфа-система подкрепления

система подкрепления, при которой веса всех активных связей c_{ij}, которые оканчиваются на некотором элементе uj, изменяются на одинаковую величину $\Delta v_{ij}(t) = \eta$, или с постоянной скоростью в течение всего времени действия подкрепления, причем веса неактивных связей за это время не изменяются.

Перцептрон, в котором используется α-система подкрепления, называется α-перцептроном.

Подкрепление называется дискретным, если величина изменения веса является фиксированной, и непрерывным, если эта величина может принимать произвольное значение.

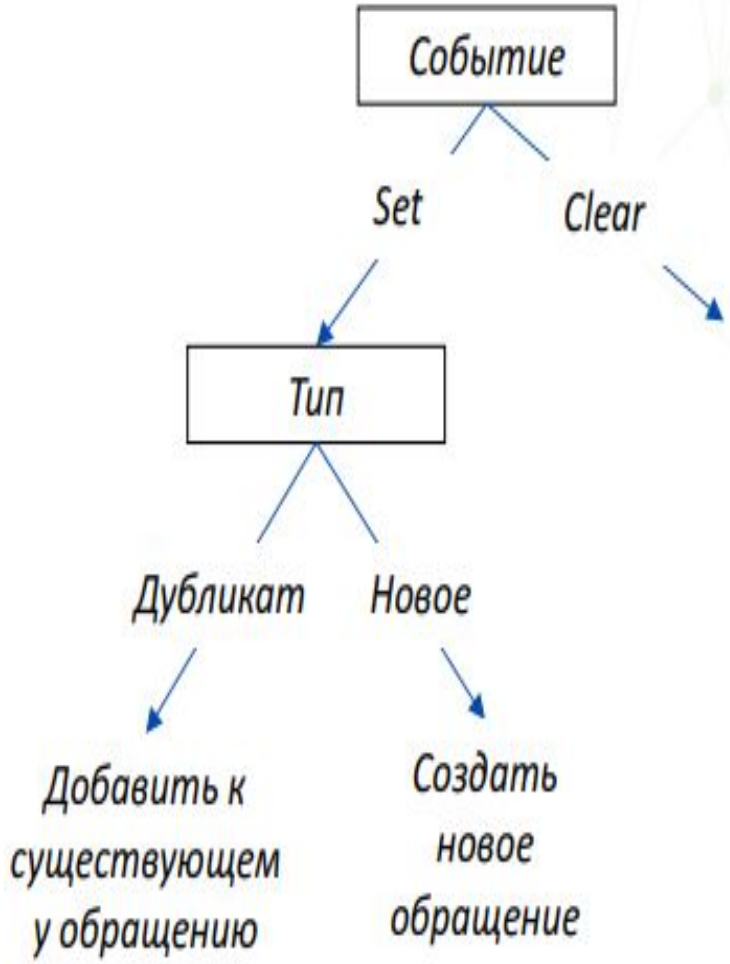
Гамма-система подкрепления

такое правило изменения весовых коэффициентов некоторого элемента, при котором веса всех активных связей сначала изменяются на равную величину, а затем из их всех весов связей вычитается другая величина, равная полному изменению весов всех активных связей, деленному на число всех связей.

Эта система обладает свойством консервативности относительно весов, так как у нее полная сумма весов всех связей не может ни возрастать, ни убывать.

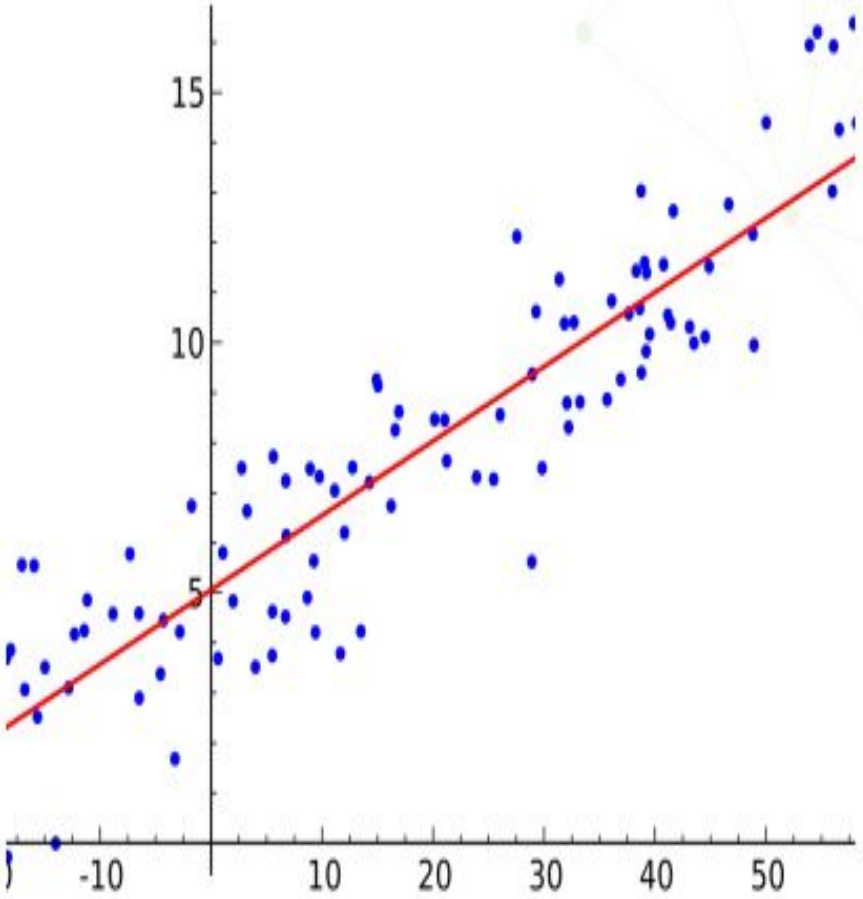
Пример: дерево принятия решений

- Граф решений и возможных действий
- Популярный алгоритм для экспертных систем



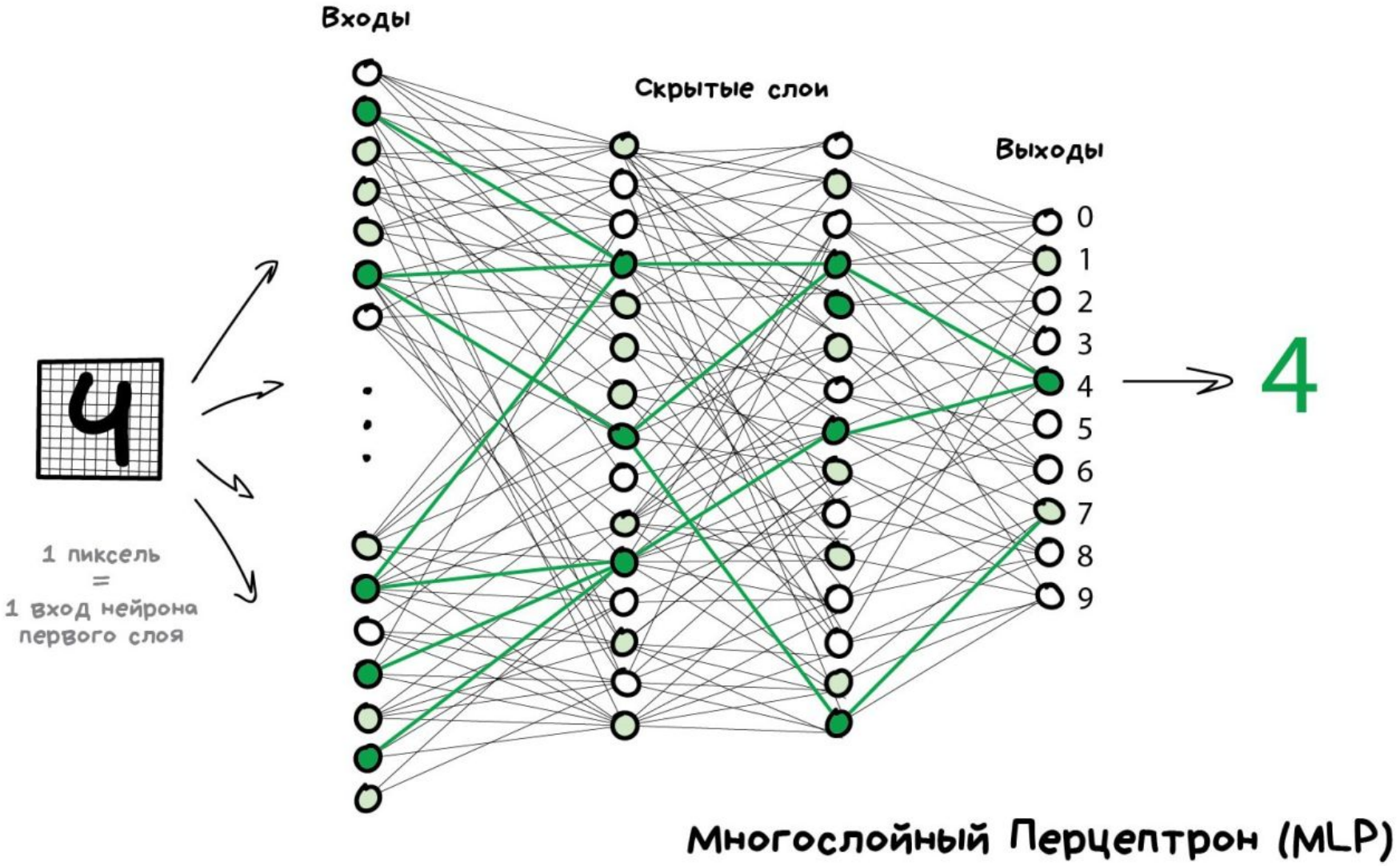
Пример: линейная регрессия

- Функция: y = **a**x+**b**
- Математический метод:
  - Решение уравнения
- Метод машинного обучения:
  - Подбор коэффициентов a и b
- Пример применения:
  - Идентификация нормального состояние загрузки канала передачи данных
  - Идентификация аномалий



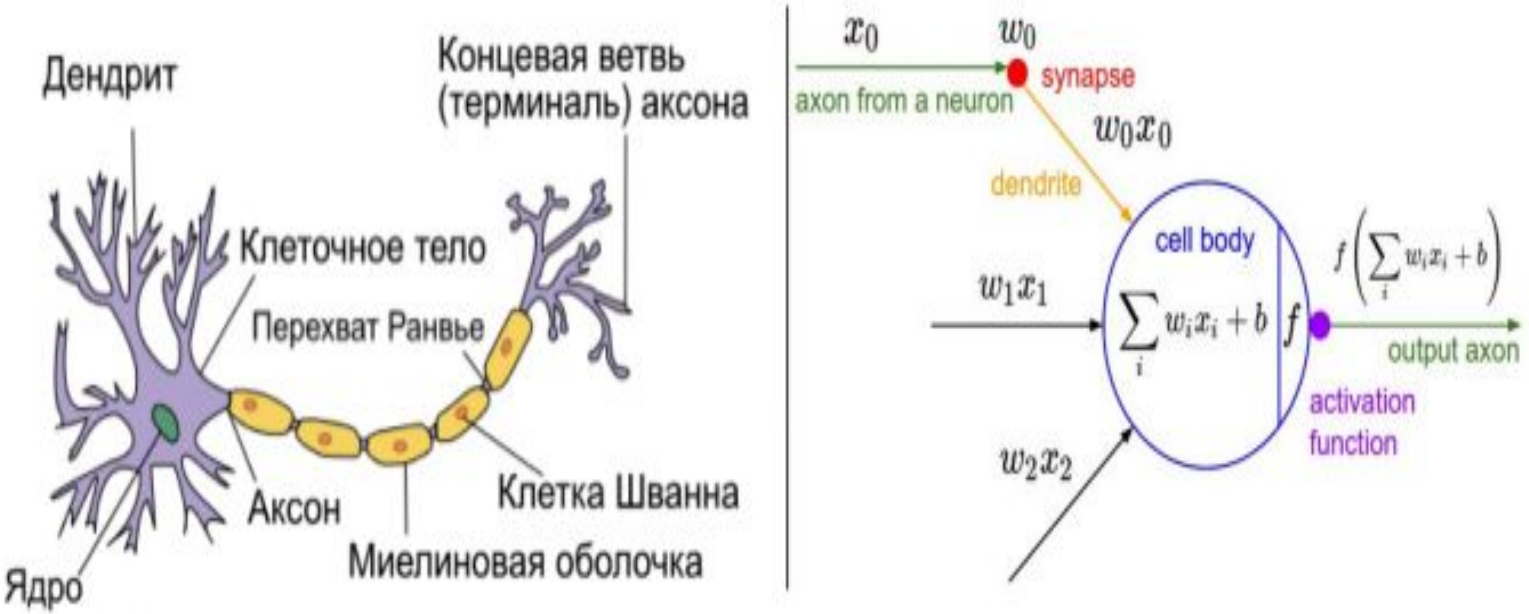
Алгоритмы машинного обучения

Нейронные сети



Нейронные сети

- Нейрон получает электрический импульс с одного или нескольких дендритов и выдает импульс через аксон



- Элемент нейронной сети выдает на выходе “1” в случае, если совокупность входных сигналов соответствует определенным значениям

1.Дерево принятия решений

Это метод поддержки принятия решений, основанный на использовании древовидного графа: модели принятия решений, которая учитывает их потенциальные последствия (с расчётом вероятности наступления того или иного события), эффективность, ресурсозатратность.

2. Наивная байесовская классификация

Наивные байесовские классификаторы относятся к семейству простых вероятностных классификаторов и берут начало из теоремы Байеса, которая применительно к данному случаю рассматривает функции как независимые.

3. Метод наименьших квадратов

Всем, кто хоть немного изучал статистику, знакомо понятие линейной регрессии. К вариантам её реализации относятся и наименьшие квадраты. Обычно с помощью линейной регрессии решают задачи по подгонке прямой, которая проходит через множество точек.

4.Логистическая регрессия

Логистическая регрессия – это способ определения зависимости между переменными, одна из которых категориально зависима, а другие независимы. Для этого применяется логистическая функция (аккумулятивное логистическое распределение). Практическое значение логистической регрессии заключается в том, что она является мощным статистическим методом предсказания событий, который включает в себя одну или несколько независимых переменных.

5. Метод опорных векторов (SVM)

Это целый набор алгоритмов, необходимых для решения задач на классификацию и регрессионный анализ. Исходя из того что объект, находящийся в N-мерном пространстве, относится к одному из двух классов, метод опорных векторов строит гиперплоскость с мерностью (N – 1), чтобы все объекты оказались в одной из двух групп.

6. Метод ансамблей

Он базируется на алгоритмах машинного обучения, генерирующих множество классификаторов и разделяющих все объекты из вновь поступающих данных на основе их усреднения или итогов голосования.

7. Алгоритмы кластеризации

Кластеризация заключается в распределении множества объектов по категориям так, чтобы в каждой категории – кластере – оказались наиболее схожие между собой элементы.

8. Метод главных компонент (PCA)

Метод главных компонент, или PCA, представляет собой статистическую операцию по ортогональному преобразованию, которая имеет своей целью перевод наблюдений за переменными, которые могут быть как-то взаимосвязаны между собой, в набор главных компонент – значений, которые линейно не коррелированы.

9. Сингулярное разложение

В линейной алгебре сингулярное разложение, или SVD, определяется как разложение прямоугольной матрицы, состоящей из комплексных или вещественных чисел.

10. Анализ независимых компонент (ICA)

Это один из статистических методов, который выявляет скрытые факторы, оказывающие влияние на случайные величины, сигналы и пр. ICA формирует порождающую модель для баз многофакторных данных.