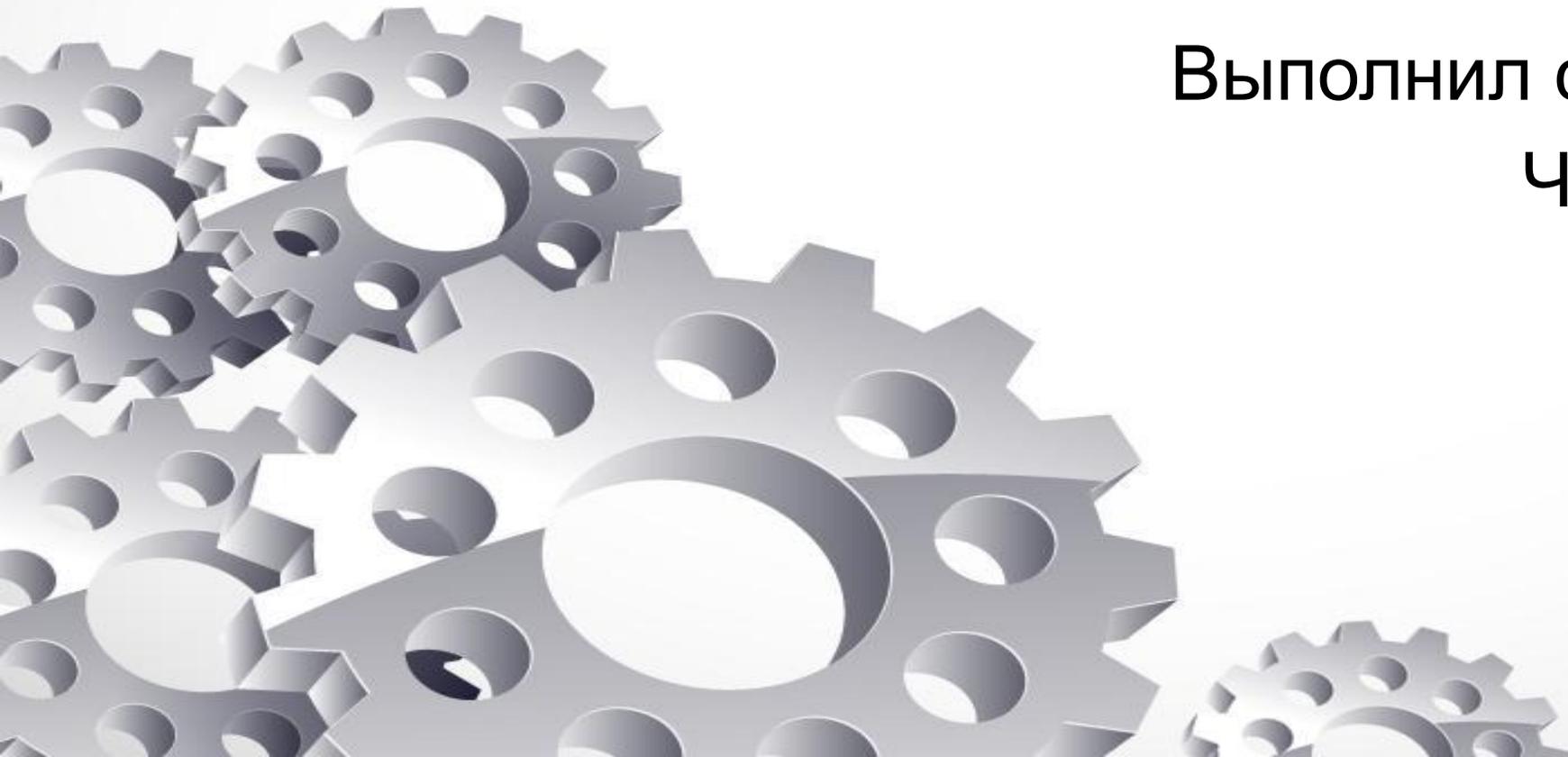
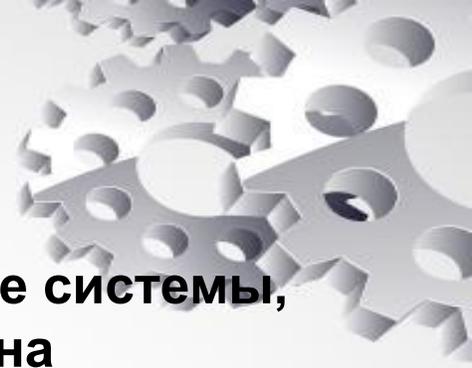


Машинное обучение

Выполнил студент 1 курса
Чауенов Камиль





- **Машинное обучение решает вопрос того, как создавать компьютерные системы, которые автоматически совершенствуются с помощью опыта. Это одна**
- **из самых быстрорастущих технических областей на сегодняшний день, лежащая**
- **на стыке информатики и статистики, а также в основе искусственного интеллекта и науки о данных. Недавний прогресс в области машинного обучения был**
- **обусловлен как разработкой новых алгоритмов и теории обучения, так и продолжающимся стремительным ростом доступности онлайн-данных и незатратных по времени вычислений. Внедрение информационно-ресурсоемких методов машинного обучения можно найти в науке, технике и торговле, что приводит**
- **к принятию решений, основанных на фактических данных, во многих сферах**
- **жизни, включая здравоохранение, производство, образование, финансовое моделирование, полицейскую деятельность и маркетинг.**



- В последнее десятилетие наблюдался быстрый рост способности сетевых
- и мобильных вычислительных систем собирать и передавать огромные объемы
- данных, явление, часто называемое “Big Data”. Ученые и инженеры,
- которые собирают такие данные, часто обращаются к машинному обучению для
- решения проблемы получения полезной информации, прогнозов и решений из
- таких наборов данных. Действительно, огромный объем данных делает необходимым
- разработку масштабируемых процедур, сочетающих вычислительные и
- статистические данные. Мобильные устройства и встроенные вычисления позволяют
- собирать большие объемы данных об отдельных людях, а алгоритмы машинного обучения
- могут извлекать уроки из этих данных, чтобы адаптировать
- свои услуги к потребностям и обстоятельствам каждого человека. Более того,
- эти персонализированные сервисы могут быть подключены, так что появляется общий
- сервис, который использует преимущества богатства и разнообразия
- данных от многих людей , в то же время адаптируясь к потребностям и обстоятельствам
- каждого.



- Наиболее широко используемыми методами машинного обучения являются **методы контролируемого обучения или обучение с учителем**. Контролируемые обучающие системы, включая классификаторы спама электронной почты, распознаватели лиц по изображениям и системы медицинской диагностики для пациентов, - все это иллюстрирует проблему аппроксимации функций где обучающие данные принимают форму набора пар (x, y) , а цель состоит в том, чтобы произвести прогноз y^* в ответ на запрос x^* .



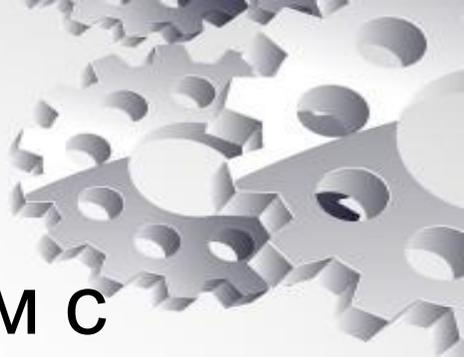


- Одной из важных областей прогресса в обучении с учителем в последние
- годы являются глубокие сети (Deep networks), которые представляют собой
- многослойные сети пороговых единиц, каждая из которых вычисляет некоторую простую параметризованную функцию своих входных данных. Системы
- глубокого обучения используют алгоритмы оптимизации на основе градиентов для настройки параметров во всей такой многослойной сети на основе ошибок на ее выходе.



- В то время как большая часть практического успеха в глубоком обучении была достигнута благодаря методам контролируемого обучения для обнаружения таких представлений, также были предприняты усилия по разработке
- алгоритмов глубокого обучения, которые обнаруживают полезные представления входных данных без необходимости в помеченных обучающих данных .
- Алгоритмы неконтролируемого обучения или обучение без учителя, вторая парадигма в исследованиях машинного обучения

- 
- Третьей важной парадигмой машинного обучения является обучение с подкреплением. Вместо обучающих примеров, которые указывают правильный результат для данного ввода, предполагается, что обучающие данные в обучении с подкреплением дают только указание на то, является ли действие правильным или нет; если действие неправильное, остается проблема поиска правильного действия.

- 
- Стоит отметить тесную связь между обучением с подкреплением и многолетними работами по обучению в психологии и неврологии, одним из заметных примеров является использование алгоритмов обучения с подкреплением для прогнозирования реакции дофаминергических нейронов у обезьян, обучающихся ассоциировать световой стимул с последующим вознаграждением сахаром



- Хотя эти три парадигмы обучения помогают систематизировать подходы, многие текущие исследования предполагают смешение этих категорий.
- Например, полуконтролируемое обучение использует немаркированные данные для дополнения помеченным данным в контексте контролируемого обучения, а дискриминационное обучение сочетает архитектуры, разработанные для неконтролируемого обучения, с оптимизационными формулировками, использующими метки. Активное обучение возникает, когда учащемуся разрешается выбирать точки данных и запрашивать у учителя целевую информацию.

Новые тенденции

- **Квантовые вычисления** - это одно из достижений, которое потенциально может расширить возможности машинного обучения.
- **Квантовые вычисления**
- позволяют выполнять одновременные операции с несколькими состояниями, обеспечивая более быструю обработку данных. В 2019 году квантовый процессор Google за 200 секунд выполнил задачу, на выполнение которой лучшему суперкомпьютеру в мире потребовалось бы 10 000 лет.





- **AutoML** стремится сделать создание приложений машинного обучения более доступным для разработчиков. Поскольку машинное обучение становится все более полезным в различных отраслях промышленности, готовые решения пользуются большим спросом. Auto-ML стремится преодолеть разрыв,
- предоставляя доступное и простое решение, которое не зависит от экспертов
- по ML.



- AutoML предлагает улучшенные инструменты маркировки данных и предоставляет возможность автоматической настройки архитектур нейронных сетей.
- Традиционно маркировка данных выполнялась вручную силами стороннего персонала. Это сопряжено с большим риском из-за человеческой ошибки. Поскольку AutoML точно автоматизирует большую часть процесса маркировки, риск
- человеческой ошибки намного ниже



- Другим примером AutoML в действии являются модели OpenAI DALL-E
- и CLIP (предварительное обучение контрастивному языковому изображению).
- Эти две модели объединяют текст и изображения для создания новых визуальных дизайнов на основе текстового описания. Одним из первых примеров
- этого в действии является то, как модели можно использовать для создания изображений на основе входного описания.

Примеры работ DALL-E



TEXT PROMPT

an armchair in the shape of an avocado. ...

AI-GENERATED
IMAGES



Edit prompt or view more images ↴

TEXT & IMAGE
PROMPT

the exact same cat on the top as a sketch on the bottom

AI-GENERATED
IMAGES



Edit prompt or view more images ↴



- Управление операционализацией машинного обучения (MLOps) - это практика разработки программных решений для машинного обучения с акцентом на надежность и эффективность. Это новый способ улучшить способ
- разработки решений для машинного обучения, чтобы сделать их более полезными для бизнеса.

- 
- GAN(Generative adversarial network) - алгоритм машинного обучения без учителя, построенный на комбинации из двух нейронных сетей, одна из которых генерирует образцы , а другая старается отличить правильные образцы от неправильных. Так как сети G и D имеют противоположные цели - создать образцы и отбраковать образцы - между ними возникает антагонистическая игра. Генеративно-состязательную сеть описал Ян Гудфеллоу из компании Google в 2014 году.



- В мире, все больше ориентированном на решения Интернета вещей, TinyML становится все более популярным. Хотя существуют крупномасштабные приложения для машинного обучения, их удобство использования довольно ограничено. Часто требуются приложения меньшего масштаба. Веб-запросу может потребоваться время для отправки данных на большой сервер, чтобы они были обработаны алгоритмом машинного обучения и затем отправлены обратно.
- Вместо этого более желательным подходом может быть использование программ ML на периферийных устройствах, что и реализуют TinyML.



- Как и в случае с любой мощной технологией, машинное обучение поднимает вопросы о том, какое из его потенциальных применений общество должно поощрять и препятствовать. Стремление в последние годы собирать новые виды персональных данных, мотивированное их экономической ценностью, приводит к очевидным проблемам конфиденциальности. Растущая ценность данных также поднимает второй этический вопрос: кто будет иметь доступ к данным и владеть ими, и кто будет пользоваться их преимуществами?
- Машинное обучение, вероятно, станет одной из самых
- преобразующих технологий 21-го века. Хотя предсказать будущее невозможно,
- важно, чтобы общество уже сейчас начало думать о том, как максимизировать свои выгоды и минимизировать риски.

СПАСИБО ЗА ВНИМАНИЕ!

