

## ТЕХНОЛОГИЧЕСКАЯ КАРТА ЗАНЯТИЯ

**Тема занятия:** Введение в машинное обучение.

**Аннотация к занятию:** обучающиеся познакомятся с базовыми понятиями машинного обучения. Узнают, как выглядят данные в машинном обучении, научатся отличать регрессию от классификации. Построят первую модель машинного обучения. В этом видео мы обсудим наиболее яркие, интересные и полезные применения машинного обучения. Поговорим о машинном обучении и его стандартной постановке.

**Цель занятия:** сформировать у учеников представление о машинном обучении, принципах его работы.

**Задачи занятия:**

- сформулировать основные понятия машинного обучения: объекты и признаки, выборку, целевую переменную;
- узнать, как выглядят данные в машинном обучении;
- познакомиться с задачами классификации и регрессии.

## Ход занятия

Этап занятия	Время	Деятельность педагога	Комментарии, рекомендации для педагогов
<b>Организационный этап</b>	5 мин.	Здравствуйте! Добро пожаловать на занятие.	Проверить готовность детей к уроку. Настроить на работу.
<b>Постановка цели и задач занятия. Мотивация учебной деятельности обучающихся</b>	7 мин.	<p>Сегодня мы познакомимся с базовыми понятиями машинного обучения.</p> <p><b>Вопросы для обсуждения:</b> Что вы можете сказать о машинном обучении своими словами? Что это такое, как вы его представляете?</p> <p><b>Возможные ответы школьников:</b> Машинное обучение позволяет компьютерам справляться с задачами, которые до недавнего времени могли решать только люди: водить автомобили, распознавать речь и переводить её, играть в шахматы и делать многое другое.</p> <p>Молодцы! На этом занятии вы узнаете, как выглядят данные в машинном обучении, и научитесь отличать регрессию от классификации. Кроме того, построите первую модель</p>	Способствовать обсуждению мотивационных вопросов.

		<p>машинного обучения. Мы обсудим наиболее яркие, интересные и полезные применения машинного обучения.</p>	
<p><b>Изучение нового материала</b></p>	<p>50 мин.</p>	<p><b>Вопрос для обсуждения</b> Давайте вспомним, что такое машинное обучение?</p> <p><b>Ответы школьников</b> Машинное обучение — это класс методов искусственного интеллекта, которые для решения задачи строят алгоритм на основе размеченных данных. То есть данных, для которых есть правильные ответы, — те, которые мы хотим научиться восстанавливать.</p> <p>Существует задача предсказания стоимости акций через определённый промежуток времени. Для её решения есть много исторических данных — курсов акций на протяжении многих лет. Эта задача часто решается именно с помощью машинного обучения.</p> <p>Другой пример — предсказание температуры воздуха, скажем, через неделю.</p> <p><b>Вопрос для обсуждения</b> Какую информацию мы можем использовать для предсказания?</p> <p><b>Ответы школьников:</b></p> <ul style="list-style-type: none"> <li>• во-первых, характеристики погоды в текущий момент времени — температура, сила и направление ветра, влажность и так далее;</li> <li>• во-вторых, исторические данные о температуре в прошлые годы.</li> </ul>	<p>Объяснение материала: беседа с использованием презентации; фронтальная работа.</p>

		<p>Отдельные элементы информации, которые можно использовать для предсказания целевого значения, называются признаками.</p> <p>В цифровом мире каждый из нас ежеминутно сталкивается с десятком алгоритмов машинного обучения, которые часто работают без вашего ведома.</p> <ul style="list-style-type: none"> <li>• Один из них — рекомендательная система. Например, любая контекстная реклама — это рекомендательная система, которая сделала предположение, что именно вам можно порекомендовать купить. Автоматическая подборка треков на музыкальной платформе, рекомендации к просмотру следующего видео, предложения интересных товаров — всё это тоже рекомендательные системы.</li> <li>• Важная и интересная область применения машинного обучения — компьютерное зрение. Например, беспилотному автомобилю, чтобы ориентироваться в пространстве, необходимо анализировать сотни изображений каждую минуту, выделяя на них важные объекты. Face ID на вашем телефоне основан на машинном обучении, сервисы для нахождения лиц и определения их возраста — продолжать можно бесконечно.</li> <li>• С помощью нейронных сетей можно рисовать новые изображения. Нейросетевая архитектура pix2pix умеет дорисовывать любой набросок до полноценного котика. А недавно Сбер выпустил новую нейросеть Kandinsky, которая рисует картины по описанию, и делает это если не на уровне Левитана и Шишкина, то весьма качественно. Вот примеры этих картин.</li> </ul>	<p>Для справки: Pix2Pix— это подход для трансляции изображений с помощью глубоких сверточ</p>
--	--	--	---

		<ul style="list-style-type: none"> <li>• Возможно, самое полезное применение машинного обучения, которого удалось добиться в последние годы — нейросеть AlphaFold, которая решает задачу предсказания пространственной структуры белка. То есть определяет, как молекула белка будет располагаться в пространстве, зная только его формулу. Картины умеют рисовать художники, предсказывать погоду неплохо получается у метеорологов, а вот предсказывать пространственную структуру белка без дорогостоящих экспериментов умеет только нейросеть. Это открыло потрясающие возможности в области синтеза различных веществ с полезными свойствами, например, инновационных лекарств.</li> </ul> <p>Научиться всему этому в рамках курса, к сожалению, невозможно.</p> <p>Искусственный интеллект — область, которая развивается очень быстро, и один человек не может охватить все её аспекты, особенно за такой короткий срок.</p> <p>Тем не менее, к концу курса некоторые из описанных применений вы научитесь самостоятельно запускать и реализовывать.</p> <p>Начнём с задачи машинного обучения.</p> <p>Для примера рассмотрим задачу оценки стоимости ноутбука. Пусть у нас есть ноутбук, стоимость которого нужно предсказать. Это поможет понять, брать его за ту или иную цену в магазине или нет. Нам известны технические характеристики этого ноутбука: количество ядер, оперативная память, объём</p>	<p>ных нейронных сетей.</p> <p>Нейросеть Kandinsky — это улучшенная версия мультимодальной ruDALL-E, генерирующей картинки по описанию.</p>
--	--	---	---

жёсткого диска, длина диагонали и время работы от аккумулятора.

**Вопрос для обсуждения**

Можете ли вы на основе этих данных предсказать стоимость ноутбука?

В предыдущих модулях мы изучали основы аналитики данных, но никакая аналитика сама по себе не сможет предсказать вам, сколько стоит данный ноутбук, если у вас не будет дополнительной информации.

**Вопрос для обсуждения**

А теперь можете? У вас появились новые данные: те же самые характеристики о дополнительных четырёх ноутбуках. Кроме того, теперь вам известна стоимость этих ноутбуков в рублях. Задача становится понятной, поскольку у вас есть, на что опираться, — данные о стоимости других ноутбуков и их сравнительные характеристики. Попробуйте, изучив характеристики ноутбуков и известные вам цены, предсказать неизвестную цену.

Правильный ответ — 86 990. Скорее всего, вы не ответили точно, но ошиблись не более, чем на 10–15 тысяч рублей. Может быть, вы не можете точно объяснить, как именно получили ответ. Так или иначе, вы проанализировали данные о четырёх ноутбуках с известной стоимостью, а затем применили полученную информацию, чтобы определить цену пятого. Возможно, вы нашли ноутбуки, наиболее близкие к нашему по характеристиками, и предположили, что стоимость будет являться чем-то средним между стоимостью этих наиболее близких ноутбуков. Именно этот процесс поиска ответа по уже

		<p>имеющимся данным и является решением стандартной задачи машинного обучения.</p> <p>Мы готовы формально описать постановку стандартной задачи машинного обучения. Пусть дано <math>X</math> большое — это наши объекты (например, множество всех ноутбуков, продающихся в магазине) и <math>Y</math> большое — множество возможных значений целевой переменной или, иначе говоря, множество ответов (например, если целевая переменная — стоимость ноутбука, то <math>Y</math> — это множество положительных чисел). Кроме того, есть закономерность <math>u</math> малое — неизвестная функция из множества <math>X</math> в множество <math>Y</math>, которая по ноутбуку определяет его настоящую стоимость.</p> <p>У нас есть и так называемая обучающая выборка — подмножество множества <math>X</math>, на котором мы знаем значения функции <math>u(x)</math>. Эти известные значения мы будем называть ответами или лейблами.</p> <p>Наша цель — подобрать алгоритм <math>u</math> с крышечкой, который приближает функцию <math>u</math> на тестовой выборке, которая недоступна в ходе подбора алгоритма. Именно это мы и делали, когда решали задачу предсказания стоимости пятого ноутбука. Здесь в роли обучающей выборки выступали те четыре ноутбука, стоимость которых нам известна. Изучив стоимости четырёх ноутбуков, мы разработали алгоритм определения стоимости, а затем применили его к пятому ноутбуку, который и является тестовой выборкой.</p> <p>Теперь попробуем выделить общую идею.</p>	
--	--	---	--

		<p>У нас есть какой-то набор данных, для которого мы знаем правильные ответы (чем бы эти ответы ни были). Этот набор данных мы отправляем в модель для ее обучения. После этого мы делаем предсказания на данных, для которых у нас еще не было правильных ответов. Наличие ответов для данных называется разметкой. Таким образом, обучающая выборка представляет из себя набор размеченных данных.</p> <p>Такая постановка задачи называется задачей обучения с учителем. Под учителем подразумеваются размеченные данные: именно они помогают модели запоминать ту или иную закономерность.</p> <p>Итак, мы познакомились с постановкой задачи машинного обучения. Правда, пока не указали, как именно эту задачу решать. Но прежде чем формально решать задачу, поговорим о том, в каком виде алгоритму подаются данные.</p> <p>Очень важно определить, что такое данные. Они могут быть чем угодно. Например, данными для машинного обучения могут быть картинки рукописных цифр. Это один из самых популярных датасетов в мире, он называется MNIST. Данные могут быть и текстом на русском языке, записью голоса или измерениями с датчика давления внутри парового котла атомной электростанции. Главное — представить эти данные в виде, пригодном для модели машинного обучения.</p> <p>Формат данных может быть разным. Самый удобный — набор или вектор из нескольких чисел. Он помогает описать каждый объект в датасете. Если строки в датасете соответствуют объектам, а столбцы — признакам, получаются стандартные табличные данные, с которыми мы уже работали.</p>	<p>Для справки: Набор MNIST — это большая коллекция рукописных цифр. Это очень популярный</p>
--	--	---	---



		<p>Если векторы признаков для разных объектов записать друг под другом в строки, то получится матрица из <math>k</math> (количества признаков) строк и <math>e</math> (количество объектов) столбцов. Эта матрица называется матрицей «объекты-признаки».</p> <p>Матрица объекты-признаки — это основной объект, который используют для обучения и применения алгоритмов машинного обучения. Она нам ещё не раз пригодится. Обратите внимание на индексирование матрицы: номера объектов обозначаются верхними индексами, а номера признаков — нижними. Скажем, в 1 строчке содержится вектор признаков первого объекта, поэтому у каждого <math>i</math> верхний индекс равен единице.</p> <p>В качестве примера давайте рассмотрим задачу классификации пассажиров «Титаника» на два класса: выжил или не выжил. Титаник — один из наиболее известных датасетов для машинного обучения. По традиции, рассмотрим его.</p> <p>Прежде чем двигаться дальше, повторим нашу терминологию. Перед вами слайд с терминами. Мы рассмотрим их на примере «Титаника».</p> <p>Во-первых, так как речь идёт об обучении с учителем, в таблице есть колонка с правильными ответами (выделена красным). Она называется целевой переменной. Все остальные колонки называются признаками (выделены синим). У нас есть 7 разнородных признаков: класс билета, имя пассажира, пол, возраст, количество дальних и ближних родственников на борту, номер билета. Каждая строка в таблице описывает один объект, то есть одного пассажира «Титаника». Все пассажиры, которые есть в таблице, составляют выборку.</p>	<p>набор данных в области обработки изображений. Он часто используется для тестирования алгоритмов машинного обучения.</p>
--	--	--	--

		<p>Мы видим, что наша таблица далека от идеала, про который мы говорили раньше. Каждая строка — это не просто набор чисел. Давайте выделим различные типы признаков, которые встречаются в этом датасете.</p> <p>Во-первых, это числовые признаки. Это наш идеал: числовые признаки легко обрабатываются любой моделью машинного обучения. В нашем датасете такие числовые признаки — это возраст, количество дальних и близких родственников.</p> <p>Категориальные признаки могут быть как строками, так и целыми числами. Обычно такие признаки имеют не очень много возможных значений. В нашем датасете категориальные признаки — это номер класса обслуживания, пол, количество дальних и близких родственников. Важно отметить, что в данном примере признаки Pclass, SibSp и Parch можно рассматривать как категориальные, несмотря на их числовую природу, потому что они принимают малое количество значений. Как обрабатывать категориальные признаки, мы с вами обсудим в одном из следующих видео.</p> <p>Наконец, бинарный признак — это категориальный признак, который принимает только два значения. В данном случае у нас только один бинарный признак — пол.</p> <p>Признаки могут не относиться ни к одному из типов, которые мы описали выше. Такой признак в нашей табличке тоже есть — это поля с именем и номером билета.</p> <p>В целом, в качестве признака может выступать совсем произвольный объект: список друзей, картинка в профиле,</p>	
--	--	--	--

запись голоса и много всего другого. Обычно для работы с такими сложными признаками применяют нейронные сети, которые могут превратить сложный признак в набор числовых признаков. Мы с вами пока не готовы к обработке таких признаков, поэтому будем их просто удалять.

Обсудим виды задач машинного обучения.

Вид задачи машинного обучения зависит от того, чем является наша целевая переменная. Так же, как и признаки, она может быть разных типов: числовая, категориальная или бинарная.

Если целевая переменная числовая, т.е. ведет себя как обычное действительное число, мы имеем дело с регрессией. Если она категориальная, т.е. имеет небольшое число возможных значений, то это задача классификации. В ней необходимо предсказать класс. Если переменная бинарная, то классификация называется бинарной.

Давайте приведём пример. Задача предсказания стоимости ноутбука — это задача регрессии, так как необходимо предсказать число — стоимость ноутбука.

Если бы нам нужно было предсказать не цену, а, скажем, офисный ноутбук или игровой, то это была бы задача классификации, причём бинарной. Точно так же задачей классификации является задача из датасета «Титаник».

Разницу между классификацией и регрессией можно показать и на графике. На первой картинке изображены объекты двух классов — красные и зелёные. Каждый объект описывается двумя признаками —  $x_1$  и  $x_2$ , и на графике такому объекту

		<p>соответствует точка с координатами <math>(x_1, x_2)</math>. Наша задача — разделить плоскость на два класса: красный класс или зелёный класс. Алгоритм в данном случае представляет из себя кривую, которая отделяет один класс от другого. Обратите внимание, что многие зелёные точки попали в группу к красным точкам, и наоборот. Это означает, что алгоритм ошибается на этих точках.</p> <p>На второй картинке изображено решение задачи регрессии. Каждый объект описывается единственным признаком <math>x</math> — координатой точки по оси абсцисс. Правильные ответы, то есть лейблы, или игреки, отложены по оси ординат. Алгоритм восстанавливает зависимость между <math>y</math> и <math>x</math> в качестве наилучшей прямой, на которую ложатся точки из выборки. Естественно, в данном случае идеальную прямую провести невозможно.</p> <p>Подведём итог. Мы изучили основные понятия машинного обучения: объекты и признаки, выборку, целевую переменную. Поговорили о том, в каком виде представляются данные в задаче машинного обучения, и обсудили задачи классификации и регрессии. На следующем уроке мы продолжим изучать понятия машинного обучения.</p>	
<p><b>Закрепление изученного материала</b></p>	<p>15 мин.</p>	<p><b>Вопросы для обсуждения</b></p> <ul style="list-style-type: none"> <li>• Сформулируйте основные понятия машинного обучения: объекты и признаки, выборку, целевую переменную.</li> <li>• Как выглядят данные в машинном обучении?</li> <li>• Что такое задачи классификации и регрессии?</li> </ul>	<p>Проверить степень усвоения учебного материала.</p>

<p><b>Этап подведения итогов занятия (рефлексия)</b></p>	<p>8 мин.</p>	<p><b>Вопросы для обсуждения:</b></p> <ul style="list-style-type: none"> <li>• Чему я научился?</li> <li>• С какими трудностями я столкнулся?</li> <li>• Каких знаний мне не хватает для более глубокого понимания изученного материала?</li> <li>• Достиг ли я поставленных целей и задач?</li> </ul>	<p>Педагог способствует размышлению обучающихся над вопросами.</p>
<p><b>Информация о домашнем задании, инструктаж по его применению</b></p>	<p>5 мин.</p>	<p>Дома повторите основные определения.</p>	

**Рекомендуемые ресурсы для дополнительного изучения:**

1. Введение в машинное обучение. [Электронный ресурс] – Режим доступа: <https://habr.com/ru/post/448892/>.
2. Нейросети Kandinsky. [Электронный ресурс] – Режим доступа: <https://habr.com/ru/news/t/671314/>.
3. Машинное обучение для начинающих. [Электронный ресурс] – Режим доступа: <https://proglib.io/p/mashinnoe-obuchenie-dlya-nachinayushchih-osnovnye-ponyatiya-zadachi-i-sfera-primeneniya-2021-08-29>.
4. Что такое искусственный интеллект. [Электронный ресурс] – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/iskusstvennyj-intellekt-voprosy-i-otvety/>.