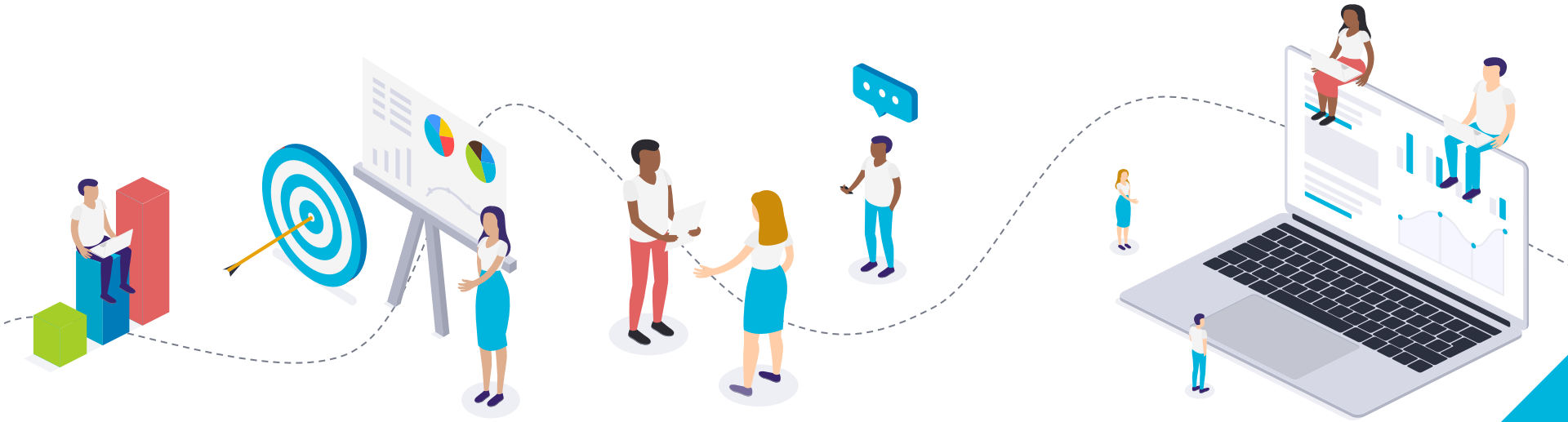
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Программа профессиональной переподготовки
«Технологии искусственного интеллекта, визуализации и анализа данных»



Что такое машинное обучение?

Машинное обучение (Machine Learning) — обширный подраздел искусственного интеллекта (ИИ), изучающий методы построения алгоритмов, способных обучаться.

Машинное обучение — это раздел ИИ, исследующий методы, позволяющие компьютерам улучшать свои характеристики на основе полученного опыта.

Вот как определяют машинное обучение представители ведущих ИТ-компаний и исследовательских центров:

Nvidia: «Это практика использования алгоритмов для анализа данных, изучения их и последующего определения или предсказания чего-либо».

Университет Стэнфорда: «Это наука о том, как заставить компьютеры работать без явного программирования».

McKinsey & Co: «Машинное обучение основано на алгоритмах, которые могут учиться на данных, не полагаясь на программирование на основе базовых правил».

Вашингтонский университет: «Алгоритмы машинного обучения могут сами понять, как выполнять важные задачи, обобщая примеры, которые у них есть».

Университет Карнеги Меллон: «Сфера машинного обучения пытается ответить на вопрос: «Как мы можем создавать компьютерные системы, которые автоматически улучшаются по мере накопления опыта и каковы фундаментальные законы, которые управляют всеми процессами обучения?»

Источник: РБК Тренды

Чем машинное обучение отличается от искусственного интеллекта (ИИ)?

ИИ подразумевает использование любого метода, который позволяет компьютеру имитировать человеческий интеллект.

Машинное обучение – разновидность ИИ, которая позволяет машине улучшать возможности решения задач по мере получения опыта.

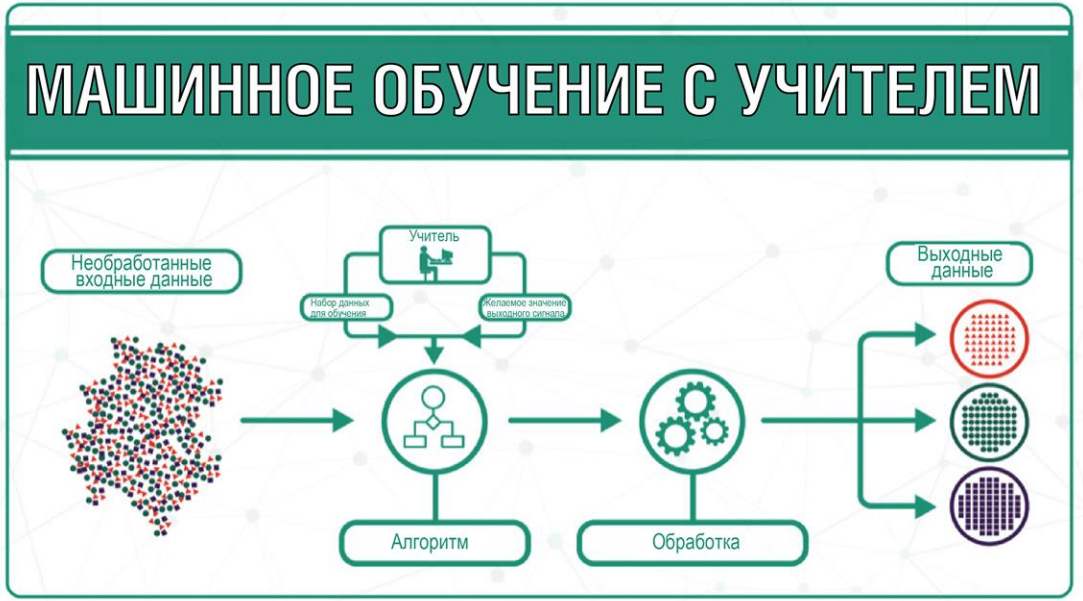
Глубокое обучение – разновидность машинного обучения, которая позволяет программному обеспечению обучаться решению задач с огромным объемом данных.

Зачем нужно машинное обучение?

- Автоматизация.
- Поиск закономерностей в данных, которые человек не сможет найти.

Обучение с учителем

Обучение с учителем — один из способов машинного обучения, в ходе которого алгоритм обучается с помощью набора примеров «объект—ответ». Процесс обучения с учителем также часто называют процессом обучения по прецедентам.



Ингредиенты машинного обучения

Объектом (x) называется то, для чего нужно сделать предсказание (абстрактная сущность).
Пространство объектов ($\boldsymbol{X}$) – это множество всех возможных объектов, для которых может потребоваться сделать предсказание.
Ответом (y) будет называться то, что нужно предсказать.
Пространство ответов ($\boldsymbol{Y}$) – множество всех возможных ответов.
Признак – это число, характеризующее объект. С помощью признаков описывают объекты.
Признаковым описанием объекта называется совокупность всех признаков: $x = (x_1, x_2, \dots, x_d)$, d – количество признаков.

признаки

	Имя	Возраст	Образование	Семейное положение	Зарплата	
объект →	Анна	25	ВО	замужем	30000	← ответ
объект →	Иван	43	СПО	женат	50000	← ответ
	Петр	35	ВО	не женат	45000	
	Василий	50	ВО	женат	60000	
	Мария	30	СПО	замужем	35000	
	Екатерина	32	ВО	не замужем	40000	

Ингредиенты машинного обучения

Цель машинного обучения состоит в применении нужных признаков объектов для построения моделей, подходящих для решения правильно поставленных задач.
Задача – это абстрактное описание проблемы, которую необходимо решить.
Под *моделью* понимают отображение исходных данных на результаты $\alpha: \boldsymbol{X} \to \boldsymbol{Y}$, которое является итогом *алгоритма* машинного обучения, примененного к *обучающим* данным

$$X = (x_i, y_i)_{i=1}^{n}$$

Следует отметить различие между задачами и проблемами обучения: задачи решаются с помощью моделей, а проблемы – алгоритмами обучения, которые порождают модели.

Имя	Возраст	Образование	Семейное положение	Зарплата
Анна	25	ВО	замужем	30000
Иван	43	СПО	женат	50000
Петр	35	ВО	не женат	45000
Василий	50	ВО	женат	60000
Мария	30	СПО	замужем	35000
Екатерина	32	ВО	не замужем	40000

Александр	30	ВО	женат	?

Ингредиенты машинного обучения

Не все алгоритмы подходят для решения определенной задачи. Поэтому вводится некоторая характеристика качества работы алгоритма — *функционал ошибки*. $Q(\alpha, X)$ — ошибка алгоритма α на выборке X.

$$Q(\alpha, X) = \frac{1}{l}\sum_{i=1}^{n}(\alpha(x_i) - y_i)^2$$ (пример функционала ошибки)

Функция, измеряющая ошибку одного предсказания, называется *функцией потерь* L. Заметим, что именно функционал качества будет определять во всех дальнейших рассуждениях, какой алгоритм является лучшим. Если метрика оценки качества выбрана неудачно и не соответствует требованиям или особенностям данных, то все дальнейшие действия обречены на провал. *Именно поэтому выбор базового функционала является крайне важным этапом в решении любой задачи анализа данных*.

Ингредиенты машинного обучения

Как только функционал качества зафиксирован, можно приступать к построению алгоритма $\alpha(x)$. Как правило, для этого фиксируют некоторое *семейство алгоритмов* $\boldsymbol{A}$, и пытаются выбрать из него алгоритм, наилучший с точки зрения функционала качества. Процесс поиска оптимального алгоритма называется *обучением*.

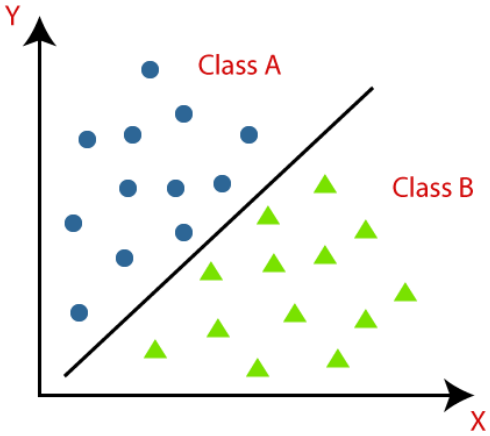
Общая постановка задачи обучения с учителем:

Для обучающей выборки нужно найти такой алгоритм $\alpha \in A$, на котором будет достигаться минимум функционала ошибки:

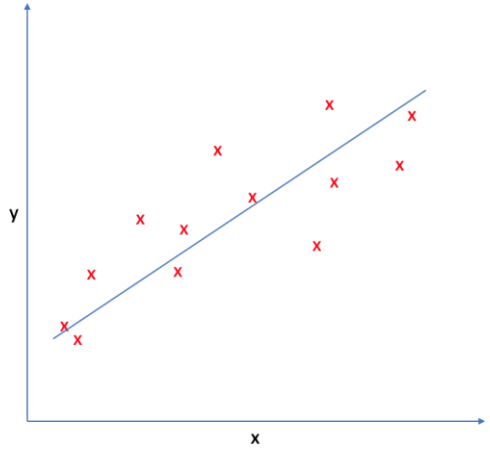
$$Q(\alpha, X) \to \min_{\alpha \in A}$$

Обучение с учителем

Задача классификации заключается в присваивании объектам категориальных, неупорядоченных значений – меток классов.



Задача регрессии заключается в предсказании значений непрерывной (вещественной) целевой переменной.



Обучение без учителя

Обучение без учителя определяет широкий класс задач обработки данных, в которых известны только признаковые описания множества объектов, и требуется обнаружить внутренние взаимосвязи, зависимости, закономерности, существующие между объектами.

Обучение без учителя

Задача кластеризации – задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров.

Задача понижения размерности заключается в том, чтобы отобразить исходные данные в пространство меньшей размерности, минимизировав при этом потери информации.

Задача поиска ассоциативных правил – задача нахождения логических закономерностей между связанными элементами.

Библиотека Scikit-learn

Scikit-learn – самая популярная библиотека для решения задач классического машинного обучения. Она предоставляет широкий выбор алгоритмов обучения с учителем и без учителя.

- Предобработка
- Классификация
- Регрессия
- Кластеризация
- Понижение размерности
- Выбор модели

https://scikit-learn.org/stable/