

Рябова Анна Сергеевна, студент 2 курса
ОГАПОУ «УАвиаК-МЦК»

Научный руководитель:
Комиссарова Диана Олеговна, преподаватель
ОГАПОУ «УАвиаК-МЦК»

ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ ДЛЯ НАЧИНАЮЩИХ

Аннотация. Статья представляет собой вводный курс в машинное обучение, предназначенный для новичков в этой области. Рассматриваются основные концепции, алгоритмы, инструменты и библиотеки, используемые в машинном обучении. Приводятся практические примеры, иллюстрирующие применение различных алгоритмов. Статья также затрагивает проблемы и задачи, решаемые машинным обучением, и предоставляет список рекомендуемой литературы для дальнейшего изучения.

Ключевые слова: Машинное обучение, алгоритмы, библиотеки, классификация, регрессия, кластеризация, Глубокие нейронные сети, обучение с учителем, обучение без учителя, обучение с подкреплением, классификация, регрессия.

Машинное обучение (Machine Learning, ML) – это подраздел искусственного интеллекта, изучающий методы построения моделей, которые способны обучаться на данных, выявлять в них закономерности и принимать решения без явного программирования на каждую конкретную задачу. Если традиционное программирование требует от человека прописывания всех возможных инструкций, то в машинном обучении компьютеру предоставляются данные и алгоритм, который самостоятельно находит решения путем комплексного статистического анализа.

Процесс обучения можно описать так: для решения задачи создается модель (математический алгоритм), которая настраивает свои параметры на основе предоставленных ей данных. Цель этого процесса – добиться, чтобы модель выдавала все более точные результаты, "приближаясь к человеческому уровню" решения аналогичной задачи. Современный бум машинного обучения стал возможен благодаря развитию глубоких нейронных сетей, содержащих тысячи и даже миллионы параметров, и росту вычислительных мощностей, особенно с появлением высокопроизводительных GPU и TPU.

Основные типы и задачи машинного обучения:

Стратегия обучения выбирается в зависимости от поставленной задачи и имеющихся данных. Выделяют три основных вида обучения.

1. Обучение с учителем (Supervised Learning)

Это обучение на размеченных примерах, где "учителем" выступают правильные ответы (метки), которые должна выдавать модель для каждого случая. Модель обучается на исторических данных, где уже известны как входные параметры, так и корректный выходной результат.

Основные задачи обучения с учителем:

1.1 Классификация: получение категориального ответа на основе набора признаков (например, "да" или "нет"). Примеры: определение наличия кота на фотографии, диагностика заболевания у пациента по медицинским показателям, фильтрация спам-писем.

1.2 Регрессия: прогноз вещественного числа на основе выборки объектов с различными признаками. Примеры: предсказание цены квартиры по ее параметрам, стоимости ценной бумаги через полгода, ожидаемого дохода магазина.

2. Обучение без учителя (Unsupervised Learning)

В этом случае модели предоставляются данные без готовых ответов или меток. Задача системы – самостоятельно найти внутренние структуры, зависимости и закономерности в данных.



Основные задачи обучения без учителя:

1.1 Кластеризация: распределение данных на группы по схожести характеристик.

Примеры: сегментация клиентов мобильного оператора по уровню платежеспособности, categorization космических объектов.

1.2 Уменьшение размерности: сведение большого числа признаков к меньшему количеству (обычно 2-3) для удобства визуализации и анализа данных.

1.3 Выявление аномалий: обнаружение редких и нестандартных случаев в данных.

Пример: выявление мошеннических операций с банковскими картами.

3. Обучение с подкреплением (Reinforcement Learning)

Это обучение методом проб и ошибок, при котором модель взаимодействует с окружающей средой и получает "поощрение" за успешные действия и "наказание" за ошибки. Такой подход используется для обучения моделей, непосредственно взаимодействующих с реальным миром: систем автоматического вождения, игровых ботов.

Популярные алгоритмы машинного обучения:

Для решения различных задач машинного обучения разработано множество алгоритмов. Вот некоторые из наиболее распространенных:

1. Линейная и полиномиальная регрессия: используются для предсказания непрерывных значений на основе исторических данных. Метод наименьших квадратов применяется для нахождения оптимальной линии (или кривой), описывающей зависимость между переменными.

2. Логистическая регрессия: хотя и называется регрессией, этот алгоритм используется для задач классификации (например, бинарной – "да/нет"). Он оценивает вероятность принадлежности объекта к определенному классу.

3. Метод опорных векторов (SVM): строит разделяющую гиперплоскость в пространстве признаков, чтобы максимизировать зазор между различными классами объектов. Эффективен для задач классификации высокой сложности.

4. Деревья решений и случайные леса: алгоритмы, которые принимают решения на основе последовательности проверок признаков объекта. Деревья просты для интерпретации, а случайный лес (ансамбль многих деревьев) повышает точность и устойчивость прогноза.

5. Метод k-ближайших соседей (kNN): классифицирует объект based on того, к какому классу принадлежит большинство из его k ближайших "соседей" в пространстве признаков.

6. Наивный байесовский классификатор: простой вероятностный классификатор, основанный на теореме Байеса и предположении о независимости признаков. Широко используется для фильтрации спама и классификации текстов.

Сферы применения машинного обучения:

Технологии машинного обучения находят применение в самых разных областях:

1. Обработка естественного языка: автоматический перевод, проверка грамматики, написание новостных статей, создание чат-ботов.

2. Компьютерное зрение: распознавание лиц, предметов и пейзажей на изображениях и видео, медицинская диагностика по снимкам.

3. Финансовая сфера: кредитный скоринг, выявление мошеннических операций, прогнозирование стоимости акций.

4. Маркетинг и ритейл: анализ успешности рекламных кампаний, сегментация клиентов, рекомендательные системы.

5. Автономные системы: управление беспилотными автомобилями, робототехника.

С чего начать изучение:

1. Курс от Михаила Коротева: включает лекции и практические задания по основам ML, регрессии, классификации, ансамблированию моделей и работе с данными. Предполагает уверенное знание Python.



2. Курс К.В. Воронцова: фундаментальный курс, читаемый в МФТИ, МГУ и Школе анализа данных Яндекса. Рассматривает математические основы ML и современные алгоритмы. Требуется знание линейной алгебры, математического анализа и теории вероятностей.

3. Курс Эндрю Ына (Andrew Ng) на Coursera: самый популярный в мире курс по основам машинного обучения от легендарного эксперта и основателя Google Brain. Служит отличным структурированным введением в тему.

Для успешного изучения машинного обучения потребуется:

1. Язык программирования Python: уверенное знание его синтаксиса и основных библиотек.

2. Математическая база: знакомство с основными понятиями линейной алгебры, математического анализа, математической статистики и теории вероятностей.

3. Популярные библиотеки и фреймворки: NumPy, Pandas, Matplotlib для анализа и визуализации данных; Scikit-learn для классических алгоритмов ML; TensorFlow или PyTorch для глубокого обучения.

Машинное обучение – это динамичная и быстро развивающаяся область на стыке математики, статистики и компьютерных наук, открывающая огромные возможности для автоматизации и решения сложных интеллектуальных задач. Начав с освоения базовых концепций, математического фундамента и инструментов (в первую очередь Python), можно постепенно переходить к решению практических задач, используя богатый выбор доступных курсов, книг и библиотек. Главное – сочетать теоретическое понимание алгоритмов с постоянной практической работой с данными.

Список литературы:

1. Босвелл, Н., Фолк, Д. Машинное обучение с использованием Scikit-learn / Н. Босвелл, Д. Фолк. – СПб.: БХВ-Петербург, 2020. – 320 с.

2. Рашка, С., Мирджалили, В. Python и машинное обучение / С. Рашка, В. Мирджалили. – М.: ДМК Пресс, 2020. – 460 с.

3. Герон, О. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. 2-е изд / О. Герон. – СПб.: БХВ-Петербург, 2020. – 816 с.

4. Бурков, А. Машинное обучение без лишних слов / А. Бурков. – М.: ДМК Пресс, 2020. – 448 с.

5. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. – М.: ДМК Пресс, 2015. – 400 с.

6. Мюллер, А.К., Гвидо, С. Введение в машинное обучение с Python / А.К. Мюллер, С. Гвидо. – М.: Альфа-книга, 2017. – 480 с.

7. Николенко, С.И., Кадури, А.А., Архангельская, Е.О. Глубокое обучение. Погружение в мир нейронных сетей / С.И. Николенко, А.А. Кадури, Е.О. Архангельская. – СПб.: Питер, 2018. – 480 с.

