

Машинное обучение. Базовое представление

Урок 5






План курса

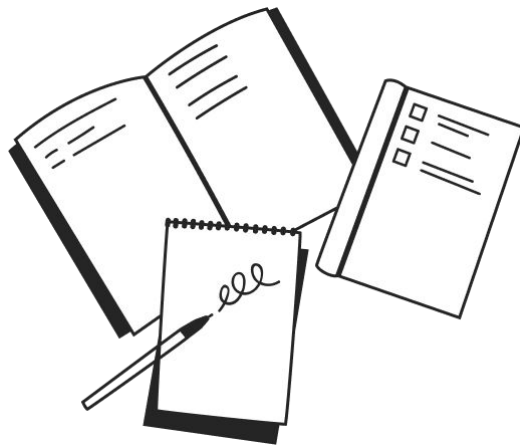
- 1 Data Science. Обзор.
- 2 Аналитика данных и ETL
- 3 Визуализация данных.

- 4 Основы статистики.
- 5 Машинное обучение. Базовое представление
- 6 Машинное обучение с учителем и без на примере классификации и кластеризации.



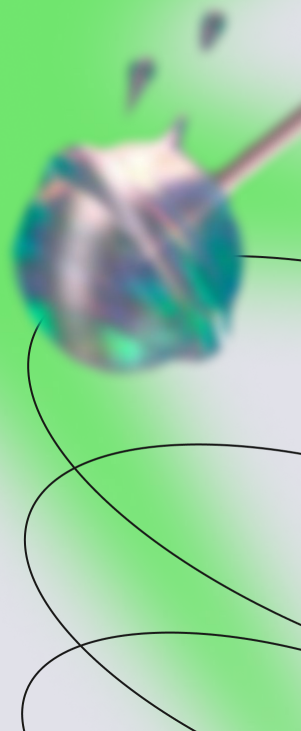
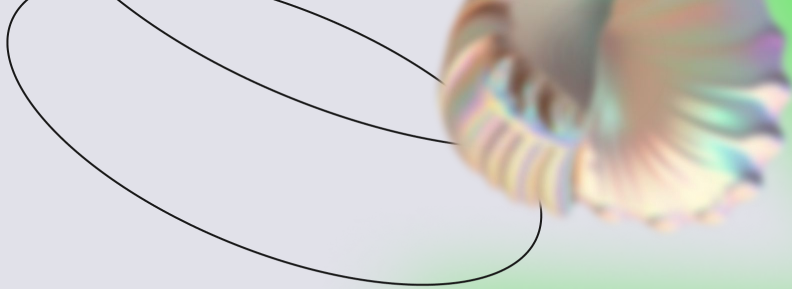
Что будет на уроке сегодня

-  Машинное обучение
-  Линейная регрессия
-  Деревья решений





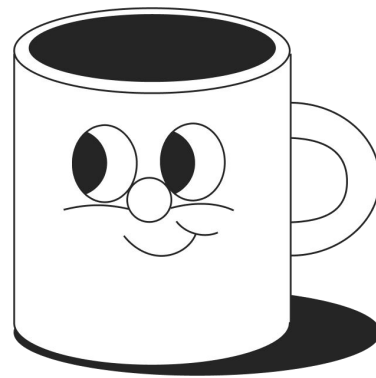
Машинное обучение





Что такое искусственный интеллект

ИИ существует как общий термин, который используется для обозначения всех компьютерных программ, способных мыслить так же, как люди. Любая компьютерная программа, демонстрирующая такие характеристики, как самосовершенствование, обучение посредством умозаключений или даже основные человеческие задачи, такие как распознавание изображений и обработка речи, считается формой ИИ.





Что такое машинное обучение

Благодаря алгоритмам машинного обучения ИИ смог выйти за рамки простого выполнения задач, для которых он был запрограммирован. До того, как машинное обучение стало мейнстримом, программы ИИ использовались только для автоматизации низкоуровневых задач в бизнесе и на предприятии.

Машинное обучение принципиально отличается от искусственного интеллекта, поскольку оно способно развиваться. Используя различные методы программирования, алгоритмы машинного обучения способны обрабатывать большие объемы данных и извлекать полезную информацию. Таким образом, они могут улучшить свои предыдущие итерации, изучая предоставленные им данные.

Сегодня алгоритмы машинного обучения обучаются с использованием трех известных методов. Вот три типа машинного обучения: обучение с учителем, обучение без учителя и обучение с подкреплением.



Типы ML



Обучение с учителем

В этом типе алгоритм машинного обучения обучается на размеченных данных



Обучение без учителя

В этом типе алгоритм машинного обучения обучается на неразмеченных данных






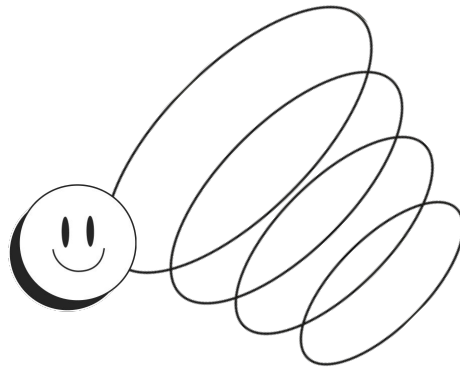
Обучение с подкреплением

Имеет алгоритм, который улучшает сам себя и учится на новых ситуациях, используя метод проб и ошибок



Приложения ML

-  Чат-боты
-  Системы рекомендаций
-  Системы интеллектуального поиска











ИИ против МЛ

Сегодня термин «искусственный интеллект» используется скорее как общий термин для обозначения технологий, обладающих когнитивными характеристиками, подобными человеческим. Как правило, исследования в области ИИ движутся к более обобщенной форме интеллекта, похожей на то, как малыши думают и воспринимают окружающий мир. Это может означать эволюцию ИИ от программы, специально созданной для одной «узкой» задачи, к решению, развернутому для «общих» решений ; вид деятельности, который мы можем ожидать от людей.

Машинное обучение, с другой стороны, является эксклюзивным подмножеством ИИ, зарезервированным только для алгоритмов, которые могут динамически улучшаться сами по себе. Они не программируются статически для одной задачи, как многие программы ИИ , и могут быть улучшены даже после их развертывания. Это не только делает их подходящими для корпоративных приложений, но и является новым способом решения проблем в постоянно меняющейся среде.



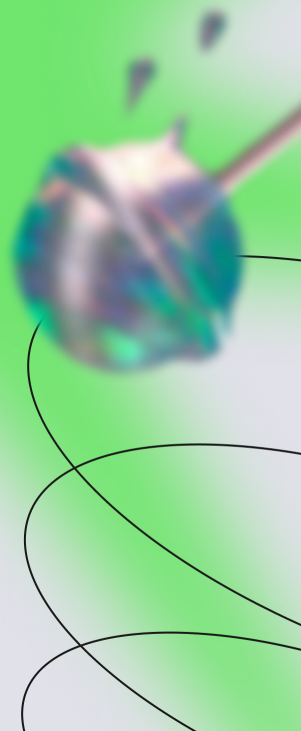
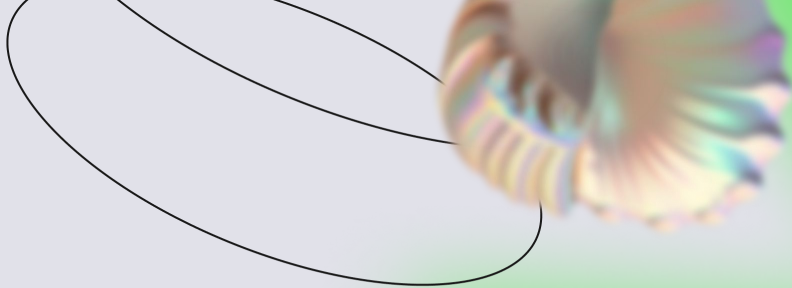
Сферы применения ML

-  Здравоохранение
-  Логистика
-  Финансы
-  Производство
-  Безопасность
-  Разведка новых месторождений





Регрессия

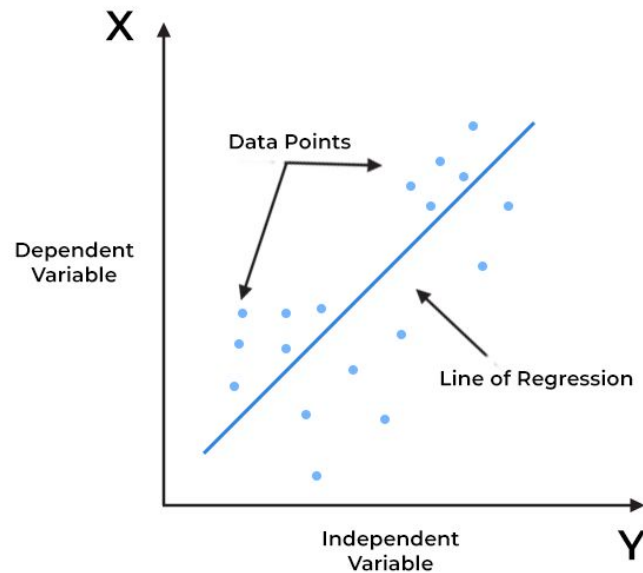




Линейная регрессия





Линейная регрессия — это алгоритм, который обеспечивает линейную связь между независимой переменной и зависимой переменной для прогнозирования исхода будущих событий. Это статистический метод, используемый в науке о данных и машинном обучении для прогнозного анализа.

Линейная регрессия — это контролируемый алгоритм обучения, который моделирует математическую связь между переменными и делает прогнозы для непрерывных или числовых переменных, таких как продажи, зарплата, возраст, цена продукта и т. д.





Преимущества

-  Простота реализации
-  Интерпретируемость
-  Масштабируемость
-  Оптимальна для онлайн-настроек

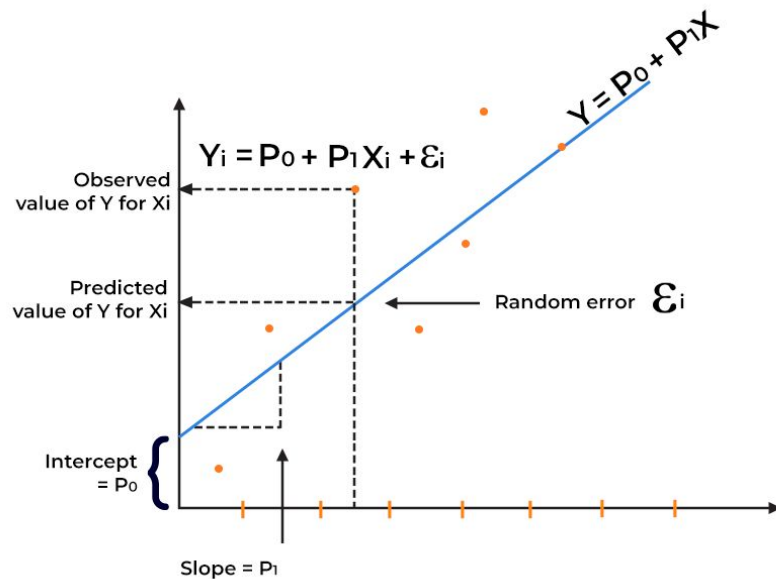


Математика

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - (mx_i + b))^2$$






Где,

- N = общее количество наблюдений (точки данных)
- $1/N \sum_{i=1}^n$ = среднее
- y_i = фактическое значение наблюдения
- $mx_i + b$ = предсказание



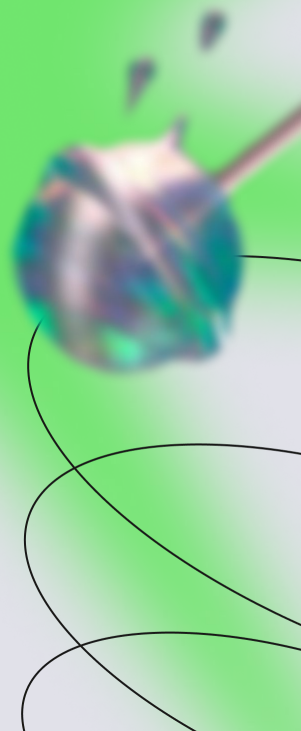
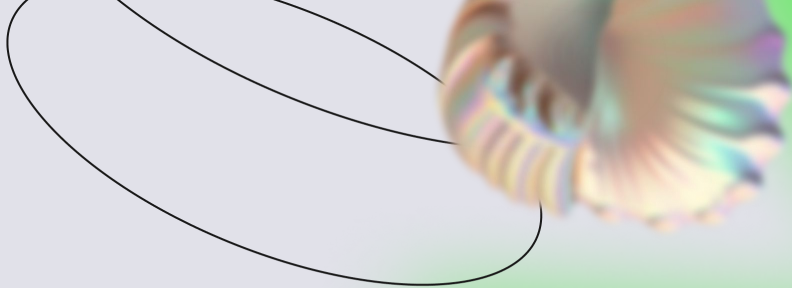


Типы линейной регрессии

-  Простая линейная регрессия
-  Множественная линейная регрессия
-  Логистическая регрессия
-  Порядковая регрессия
-  Полиномиальная логистическая регрессия



Дерево решений

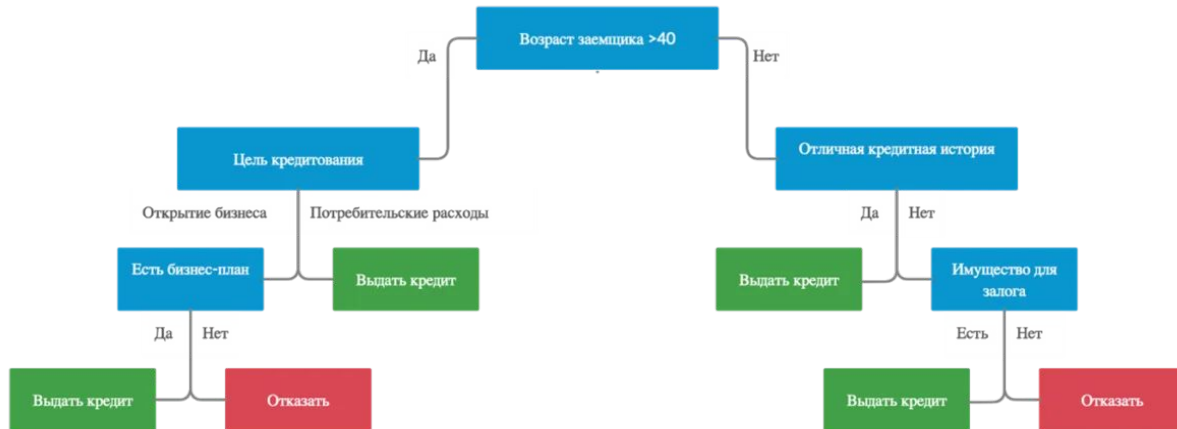




Деревья решений







Деревья решений — это, по сути, схематические подходы к решению проблем.

Дерево решений — метод представления решающих правил в определенной иерархии, включающей в себя элементы двух типов — узлов (node) и листьев (leaf). Узлы включают в себя решающие правила и производят проверку примеров на соответствие выбранного атрибута обучающего множества.









Преимущества

-  Формируют четкие и понятные правила классификации
-  Способны генерировать правила в областях, где специалисту трудно формализовать свои знания
-  Легко визуализируются
-  Быстро обучаются и прогнозируют
-  Не требуется много параметров модели
-  Поддерживают как числовые, так и категориальные признаки








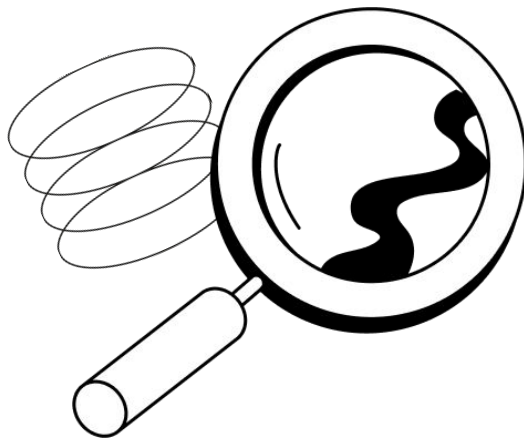
Недостатки

-  Деревья решений чувствительны к шумам во входных данных
-  Разделяющая граница имеет определенные ограничения
-  Возможно переобучение
-  Сложный поиск оптимального дерева решений



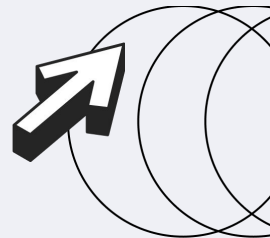
Где применяются деревья решений

-  Банковское дело
-  Промышленность
-  Медицина
-  Молекулярная биология
-  Торговля





Давайте делать это в коде!





Спасибо за внимание

