

A night city street scene with a tram and a digital data overlay. The street is illuminated by streetlights, and a tram is visible in the distance. A large, glowing digital data overlay consisting of binary code (0s and 1s) is superimposed on the right side of the image, extending from the foreground into the background. The overlay is composed of multiple lines of binary code, with some characters appearing larger and more prominent than others. The overall color palette is dominated by blues and yellows, with the digital overlay adding a vibrant, futuristic feel.

Программное обеспечение учетных и статистических операций

Введение в бизнес-анализ

- Машинное обучение
- Обучающая выборка

Введение в алгоритмы DM

Машинное обучение

Машинное обучение (*Machine Learning*) - обширный подраздел искусственного интеллекта, изучающий методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения;

- находится на стыке математической статистики, методов оптимизации и дискретной математики, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения;
- методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам;
- методы тесно связаны с извлечением информации (*Information Extraction*), интеллектуальным анализом данных (*Data Mining*).

Введение в алгоритмы DM

Машинное обучение. Основные понятия

Обучаемая модель - обширный подраздел искусственного интеллекта, изучающий методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения;

Обучающая выборка - обширный подраздел искусственного интеллекта, изучающий методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения;

Обучение - обширный подраздел искусственного интеллекта, изучающий методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения;

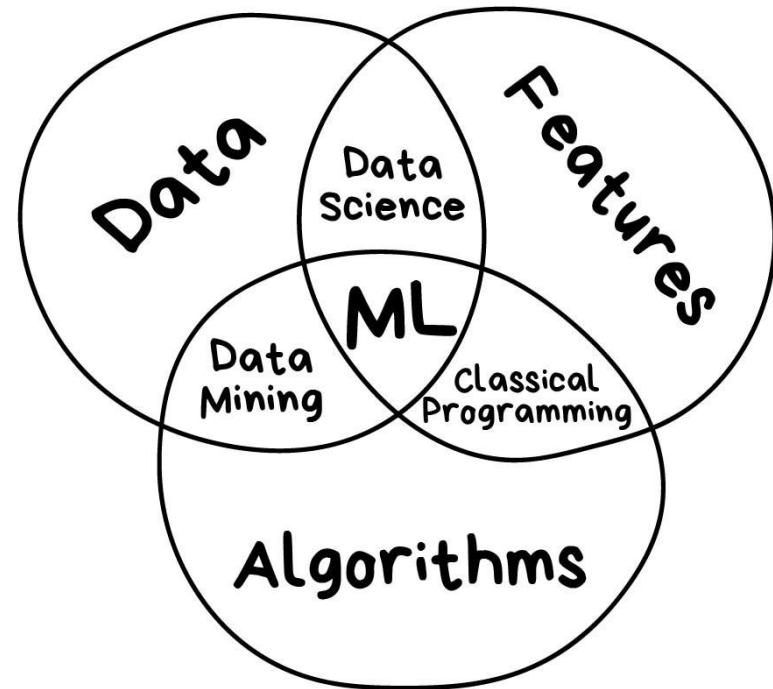
Алгоритм обучение - обширный подраздел искусственного интеллекта, изучающий методы построения моделей, способных обучаться, и алгоритмов для их построения и обучения;

Введение в алгоритмы DM

Цель машинного обучения — предсказать результат по входным данным. Чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат

Три составляющие ,
необходимые для обучения:

- ✓ ДАННЫЕ
- ✓ ПРИЗНАКИ
- ✓ АЛГОРИТМ



Введение в алгоритмы DM



- **Искусственный интеллект** — название всей области, как биология или химия.
- **Машинное обучение** — это раздел искусственного интеллекта. Важный, но не единственный.

- **Нейросети** — один из видов машинного обучения. Популярный, но есть и другие, не хуже.
- **Глубокое обучение** — архитектура нейросетей, один из подходов к их построению и обучению. На практике сегодня мало кто отличает, где глубокие нейросети, а где не очень. Говорят название конкретной сети и всё.

Введение в алгоритмы DM

Машина может

Предсказывать

Запоминать

Воспроизводить

Выбирать лучшее

Машина не может

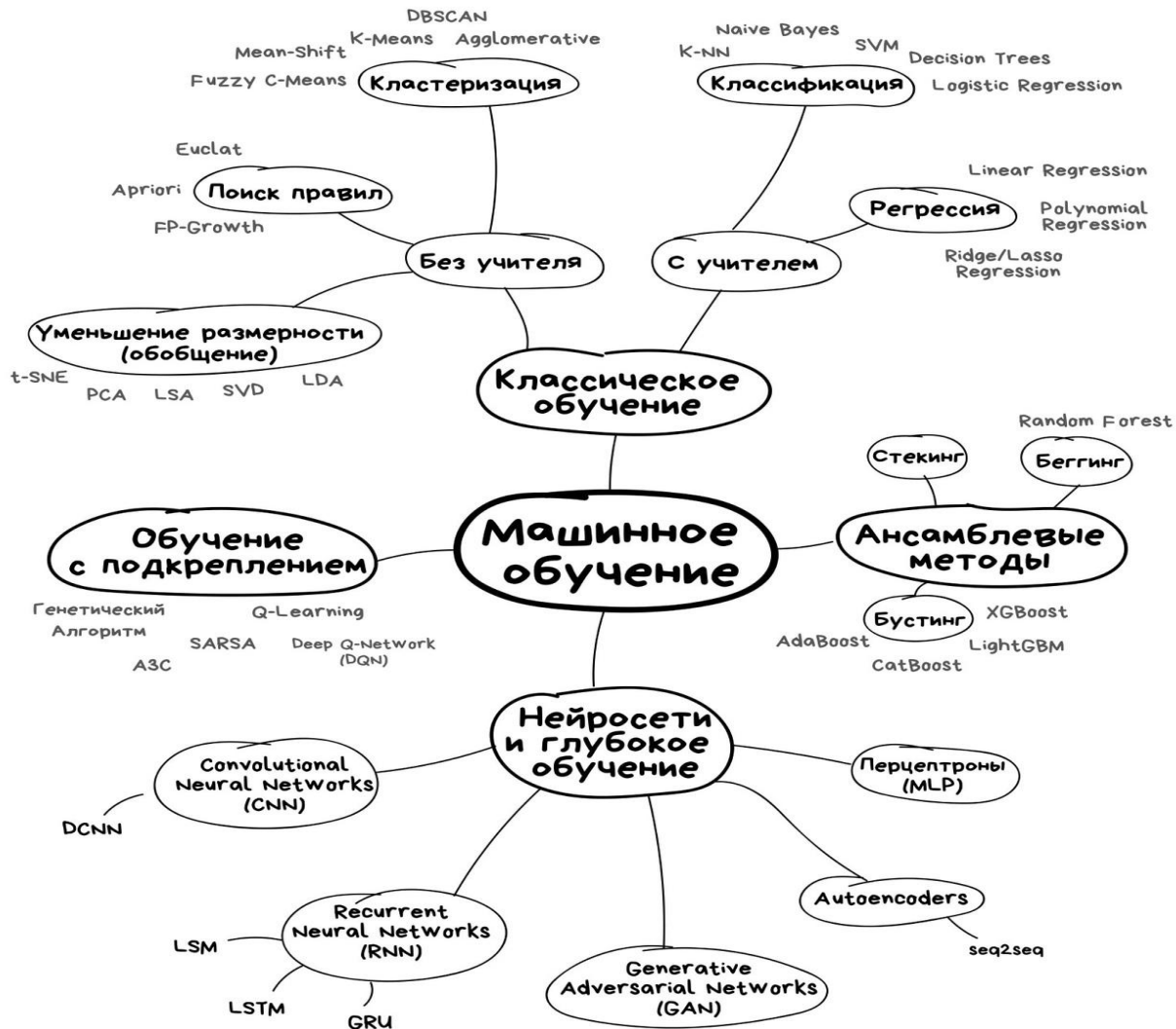
Создавать новое

Резко поумнеть

Выйти за рамки задачи

Убить всех людей

Введение в алгоритмы DM



Введение в алгоритмы DM

В машинном обучении есть всего четыре основных направления

Основные виды машинного обучения



Введение в алгоритмы DM



Введение в алгоритмы DM

С учителем машина обучится быстрее и точнее, потому в боевых задачах его используют намного чаще.

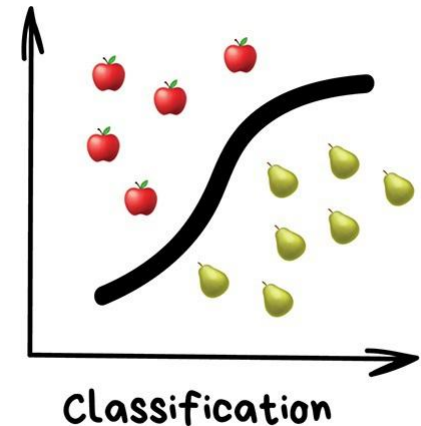
Задачи делятся на два типа:

- ❑ **классификация** — предсказание категории объекта;
- ❑ **регрессия** — предсказание места на числовой прямой.

Сегодня используют для:

- ✓ Спам-фильтры
- ✓ Определение языка
- ✓ Поиск похожих документов
- ✓ Анализ тональности
- ✓ Распознавание рукописных букв и цифр
- ✓ Определение подозрительных транзакций

Популярные алгоритмы: Наивный Байес, Деревья Решений, Логистическая Регрессия, К-ближайших соседей, Машины Опорных Векторов



Введение в алгоритмы DM

!!! Классификация вещей — самая популярная задача во всём машинном обучении

привет...	1829
валера ...	1710
нет ...	1191
куда ...	1012
небо ...	985
огурцы ...	873
говорить...	747
третий ...	739

нормальные
письма

виагра ...	1552
казино ...	1492
100% ...	1320
кредит...	1184
скидка ...	985
нажми ...	873
free ...	747
доход ...	739

спам-письма

672 раза

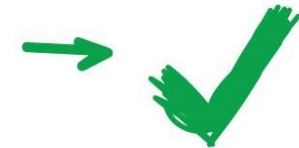
«КОТИК»

13 раз

Простейший спам-фильтр
(использовались года до 2010)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

формула Байеса



не спам

Наивный Байес

Введение в алгоритмы DM

!!! Классификация вещей — самая популярная задача во всём машинном обучении

Давать ли кредит?



Дер

Введение в алгоритмы DM

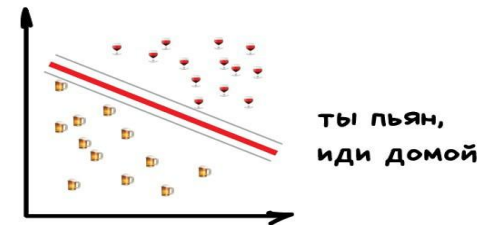
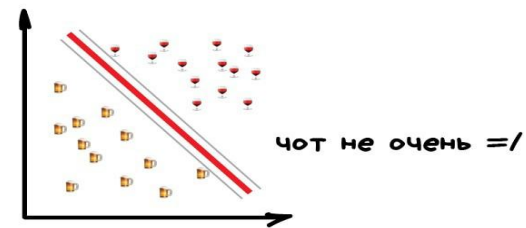
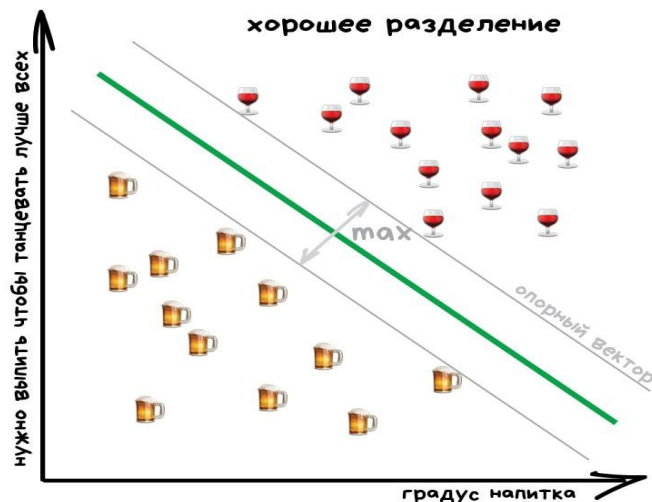
!!! Классификация вещей — самая популярная задача во всём машинном обучении

Самым популярным методом классической классификации заслуженно является **Метод Опорных Векторов**.

Классифицировали: виды растений, лица на фотографиях, документы по тематикам и пр.

Ищем, как так провести две прямые между категориями, чтобы между ними образовался наибольший зазор

Разделяем виды алкоголя



Метод Опорных Векторов

Введение в алгоритмы DM

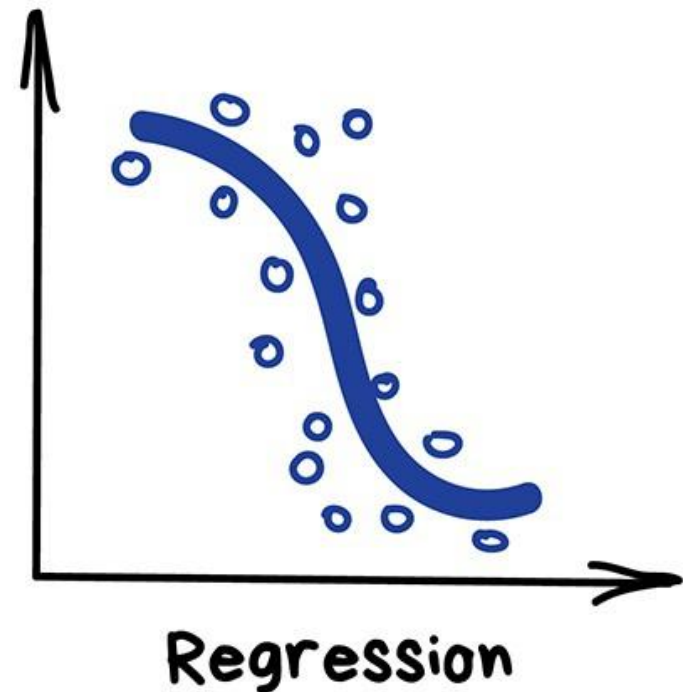
Регрессия — та же классификация, только вместо категории предсказываем число.
На регрессию идеально ложатся любые задачи, где есть зависимость от времени.

Сегодня используют для:

- ✓ Прогноз стоимости ценных бумаг
- ✓ Анализ спроса, объема продаж
- ✓ Медицинские диагнозы

Любые зависимости числа от времени

Популярные алгоритмы: Линейная или
Полиномиальная Регрессия



Введение в алгоритмы DM

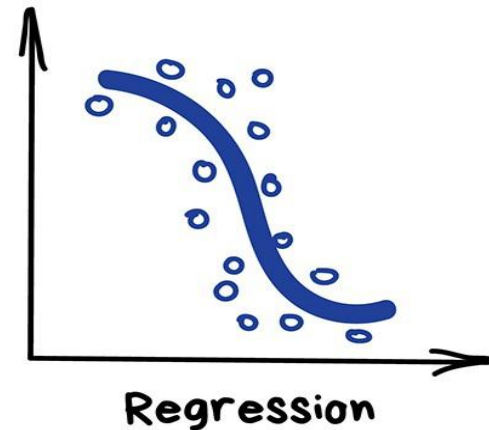
Регрессия — та же классификация, только вместо категории предсказываем число.
На регрессию идеально ложатся любые задачи, где есть зависимость от времени.

Сегодня используют для:

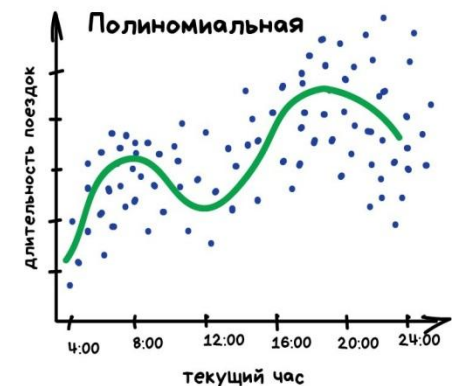
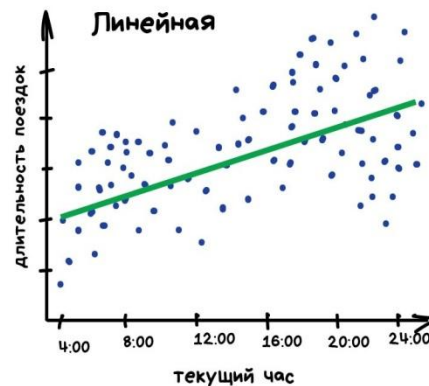
- ✓ Прогноз стоимости ценных бумаг
- ✓ Анализ спроса, объема продаж
- ✓ Медицинские диагнозы

Любые зависимости числа от времени

Популярные алгоритмы: Линейная или
Полиномиальная Регрессия



Предсказываем пробки



Регрессия

Введение в алгоритмы DM

Обучение без учителя было изобретено позже и на практике используется реже

