

МАШИННОЕ ОБУЧЕНИЕ



Проскуряков Родион Александрович, 2019

ЧТО ЭТО ТАКОЕ?

Машинное обучение - это область науки, занимающаяся исследованием алгоритмов и статистических моделей, которые компьютеры используют для выполнения конкретных задач, без четких инструкций. Алгоритмы машинного обучения опираются на данные, в которых они самостоятельно ищут закономерности.



ЗАЧЕМ ЭТО ВСЕ НУЖНО?

Хотим имея какие-то данные, научить машину делать предсказания, т. е. чтобы машина нашла **зависимость**.

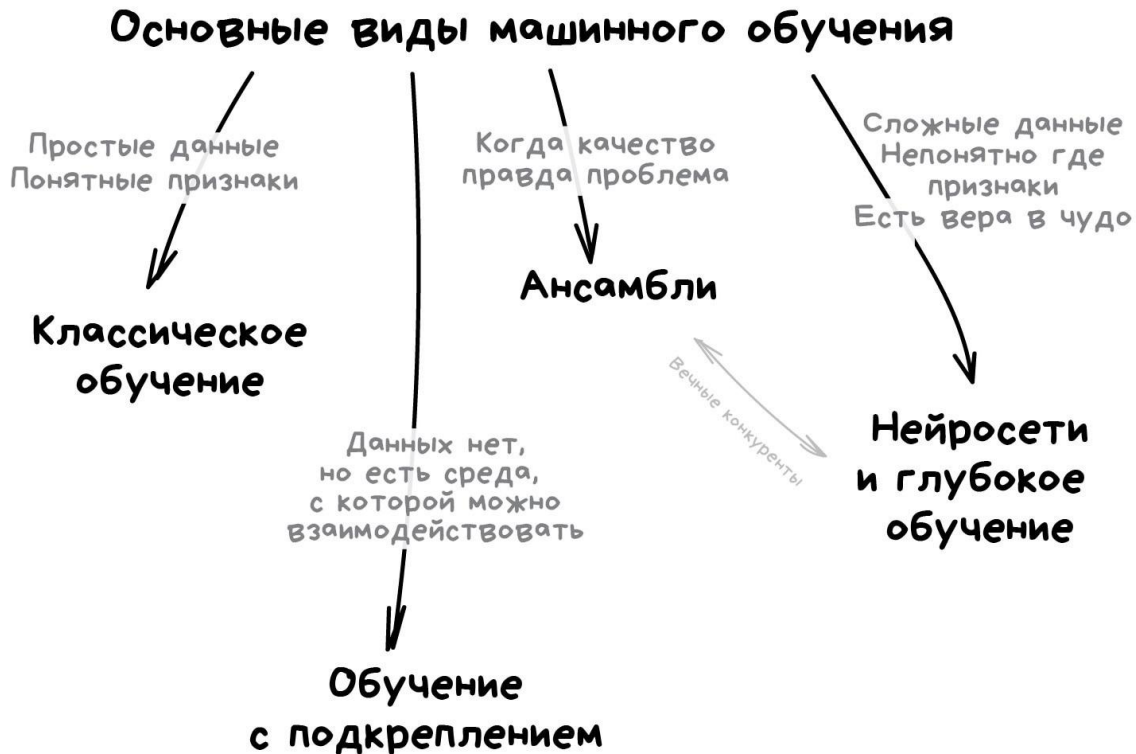
Знаем, как что в часе 60 минут: $M = 60H$

А как зависит цена квартиры от ее площади, числа комнат, расположения относительно магазинов и т. д.?

А ЧТО ПРО ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ?



ОСНОВНЫЕ ВИДЫ МАШИННОГО ОБУЧЕНИЯ



Как машины ведут себя при пожаре

Классическое программирование

«Я просчитал все варианты
событий и ты сейчас
должен связать верёвку
из хлебного мякиша»

Машинное обучение

«По статистике люди гибнут
в 6% пожаров, поэтому
рекомендую вам умереть
прямо сейчас»

Обучение с подкреплением

«Да просто беги от огня
AAAAAAAAAAAA!!!!
»

КЛАССИЧЕСКОЕ ОБУЧЕНИЕ

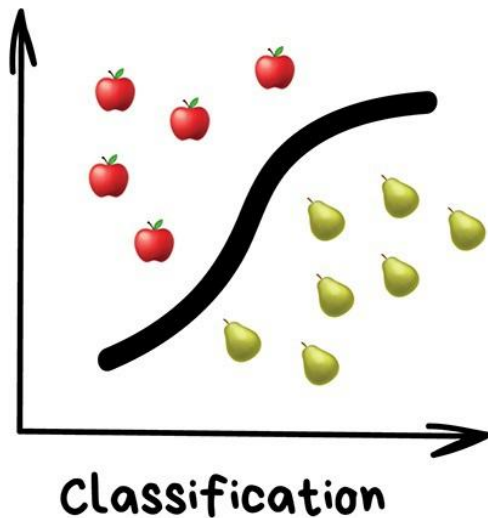


КЛАССИФИКАЦИЯ

«Разделяет объекты по заранее известному признаку. Носки по цветам, документы по языкам, музыку по жанрам»

Сегодня используют для:

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



ДЕРЕВЬЯ РЕШЕНИЙ

Давать ли кредит?



Дер

НАИВНЫЙ БАЙЕС

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

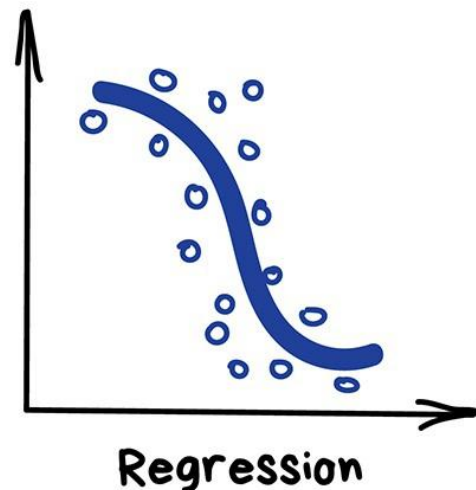
Наивный Байес

ЗАДАЧА РЕГРЕССИИ

«Нарисуй линию вдоль моих точек. Да, это машинное обучение»

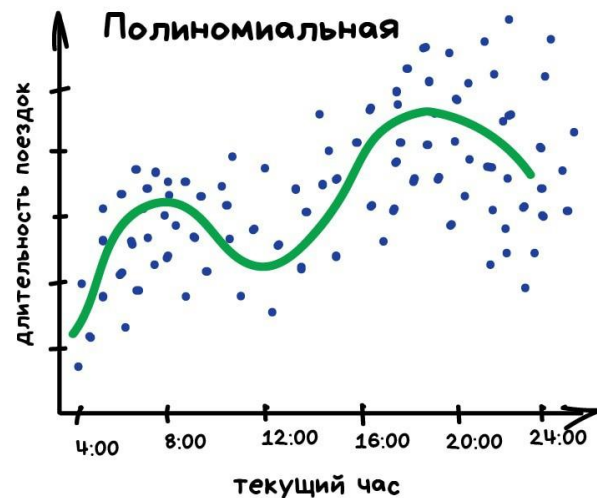
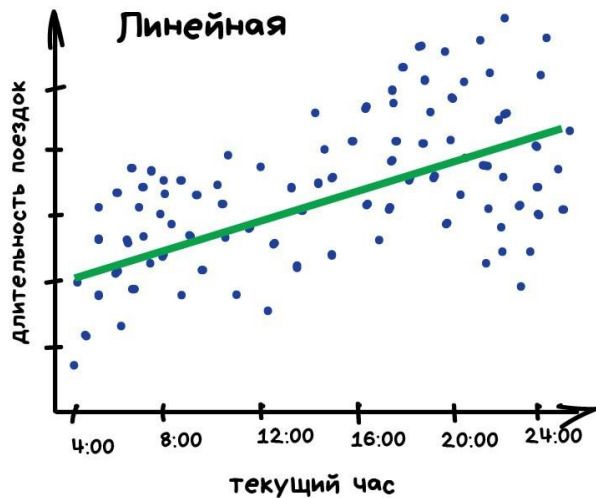
Сегодня используют для:

- Прогноз стоимости ценных бумаг
- Анализ спроса, объема продаж
- Медицинские диагнозы
- Любые зависимости числа от времени



ПРИМЕР ЗАДАЧИ РЕГРЕССИИ

Предсказываем пробки



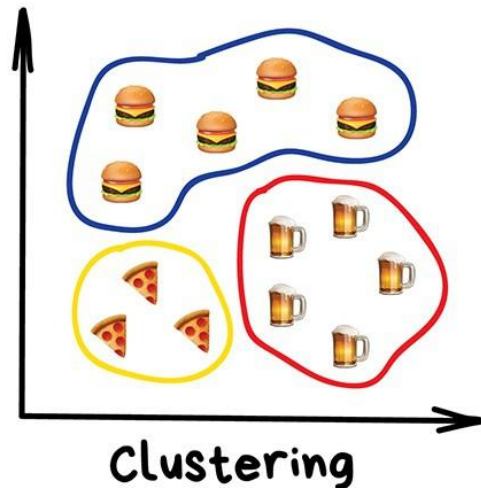
Регрессия

ЗАДАЧА КЛАСТЕРИЗАЦИИ

«Разделяет объекты по неизвестному признаку. Машина сама решает как лучше»

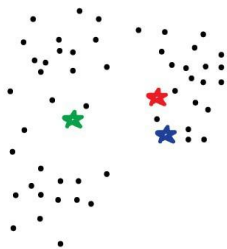
Сегодня используют для:

- Сегментация рынка (типов покупателей, лояльности)
- Объединение близких точек на карте
- Сжатие изображений
- Анализ и разметка новых данных
- Детекторы аномального поведения

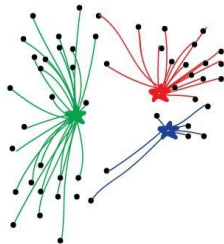


КЛАСТЕРИЗАЦИЯ

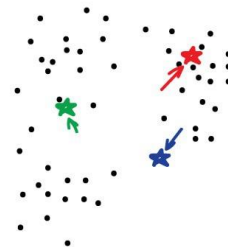
Ставим три ларька с шаурмой оптимальным образом
(иллюстрируя метод K-средних)



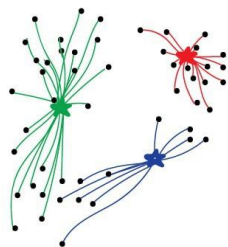
1. Ставим ларьки с шаурмой в случайных местах



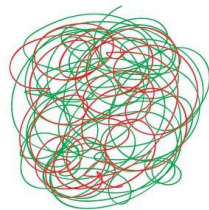
2. Смотрим в какой кому ближе идти



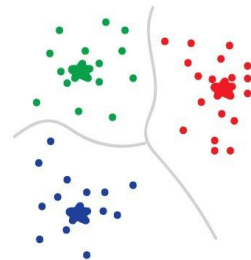
3. Двигаем ларьки ближе к центрам их популярности



4. Снова смотрим и двигаем



5. Повторяем много раз



6. Готово, вы великолепны!

ЧТО НУЖНО ЗНАТЬ, ДЛЯ ТОГО, ЧТОБЫ ЗАНИМАТЬСЯ ML?

1. Python

[Курс на stepik \(если вы раньше не программировали\)](#)

[Курс на stepik \(если уже знакомы с другим языком\)](#)

2. Немного математики

3. Классические алгоритмы (этим мы и будем заниматься вначале)

4. Нейронные сети? (это будет попозже)