

МАШИННОЕ ОБУЧЕНИЕ

МАШИННОЕ ОБУЧЕНИЕ

МАШИННОЕ ОБУЧЕНИЕ

МАШИННОЕ ОБУЧЕНИЕ

КАК ЗАСТАВИТЬ МАШИНУ ЧТО-ТО ВЫПОЛНИТЬ?

Зайди на сайт

Отсортируй числа по возрастанию

Если файл очень большой, положи его в
архив

**СТАНДАРТНЫЕ АЛГОРИТМЫ РАБОТАЮТ,
КОГДА НЕТ НЕОДНОЗНАЧНОСТИ**

**ИЗНАЧАЛЬНО ВСЯ ИНФОРМАЦИЯ
ДЛЯ КОМПЬЮТЕРА ПРЕДСТАВЛЕНА ПРИМЕРНО ТАК:**

```
000001101001010010011010001010100101010010100101001010010  
1111100001101001010010100101010101010101011111110010  
010100101101101010101010101010100111010010101001010  
1001010101001111010101001010001101111110100101010010101  
00101010010101101111111010010010100101001010100101001  
01010100101010100101010010100101001010010111111010010...
```

МЫ МОЖЕМ СКАЗАТЬ:

0001 — это буква А

0010 — это буква Б

...

А ЗАТЕМ:

«Если слово кончается на ая, ое или ый» — это
прилагательное. Но есть исключения: ...

**МЫ МОЖЕМ РАБОТАТЬ С ОБЪЕКТАМИ
ТОГДА, КОГДА
ОПИШЕМ КОМПЬЮТЕРУ НА ЕГО ЯЗЫКЕ,
ЧТО ОНИ ОЗНАЧАЮТ**

**НО ЧТО, ЕСЛИ ОБЪЕКТЫ СЛОЖНО ОПИСАТЬ
НА ЯЗЫКЕ ЦИФР?**

**КАК БЫ ВЫ ОПИСАЛИ
КОТИКА?**



ИЛИ УДИВЛЕНИЕ?



МАШИННОЕ ОБУЧЕНИЕ РАБОТАЕТ ПРИМЕРНО ТАК:

Смотри, это кот:



МАШИННОЕ ОБУЧЕНИЕ РАБОТАЕТ ПРИМЕРНО ТАК:

И это кот:



МАШИННОЕ ОБУЧЕНИЕ РАБОТАЕТ ПРИМЕРНО ТАК:

А это не кот:



(и так желательно пару тысяч раз)

После этого компьютер сам выводит какие-то правила, по которым может отличить кота от всего остального

Этот алгоритм называется классификация

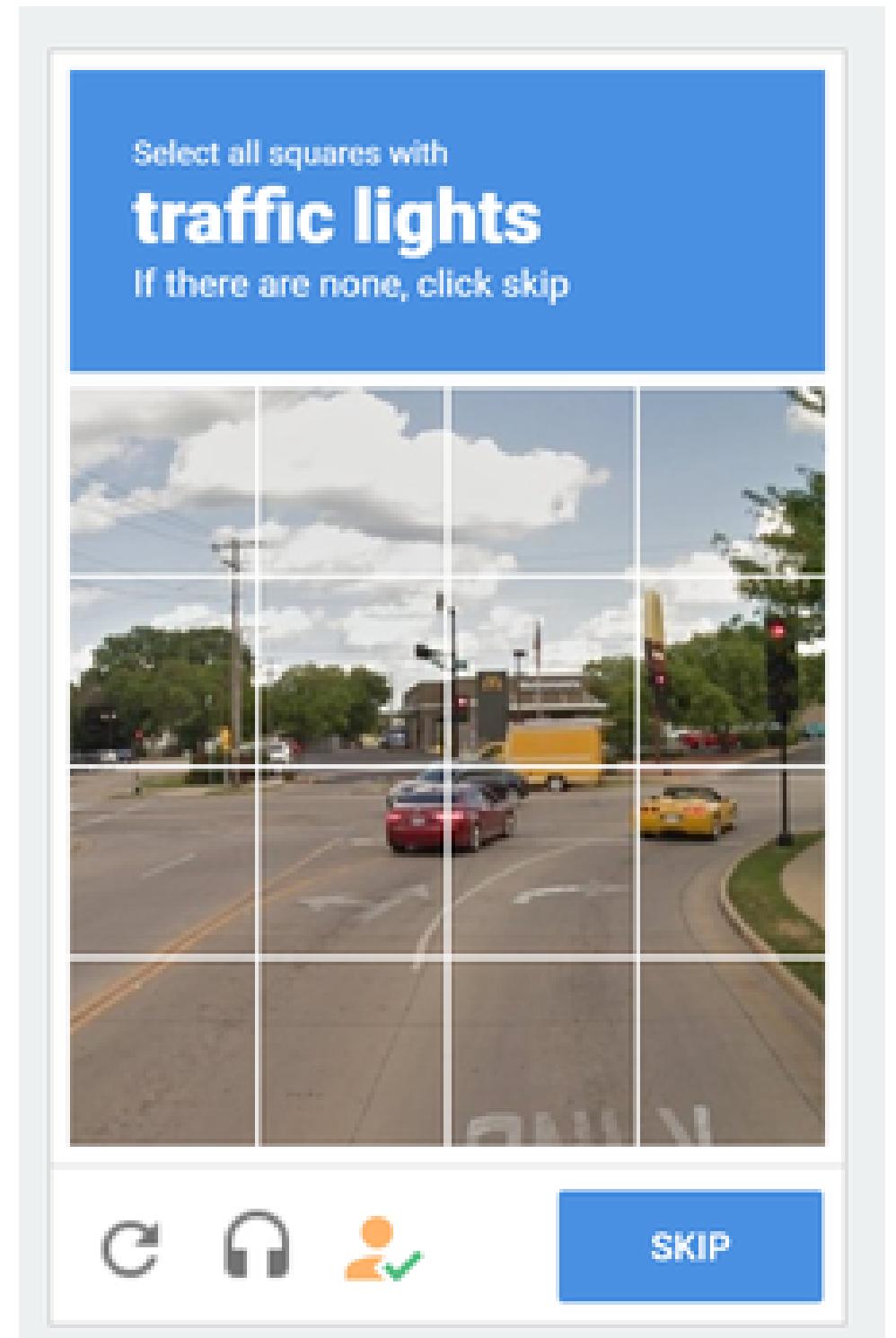
**За алгоритмами стоит сложная математика, но
её не обязательно понимать, чтобы успешно их
применять:**

$$\begin{aligned}\log P(\vec{y} \mid X, \vec{w}) &= \log \prod_{i=1}^{\ell} P(y = y_i \mid \vec{x}_i, \vec{w}) \\&= \log \prod_{i=1}^{\ell} \sigma(y_i \vec{w}^T \vec{x}_i) \\&= \sum_{i=1}^{\ell} \log \sigma(y_i \vec{w}^T \vec{x}_i) \\&= \sum_{i=1}^{\ell} \log \frac{1}{1 + \exp^{-y_i \vec{w}^T \vec{x}_i}} \\&= - \sum_{i=1}^{\ell} \log(1 + \exp^{-y_i \vec{w}^T \vec{x}_i})\end{aligned}$$

**МАШИННОЕ ОБУЧЕНИЕ РЕШАЕТ РАЗНЫЕ ЗАДАЧИ,
И НЕ КАЖДЫЙ АЛГОРИТМ ОБУЧАЕТСЯ НА
ПРИМЕРАХ**

**НО ЧАЩЕ ВСЕГО НАМ НУЖНО МНОГО
РАЗМЕЧЕННЫХ ДАННЫХ**

**СОЗДАВАТЬ ТАКИЕ ДАННЫЕ –
БОЛЬШОЙ ТРУД**





Источник: https://vas3k.ru/blog/machine_learning/

АЛГОРИТМЫ КЛАССИФИКАЦИИ В ЖУРНАЛИСТИКЕ

Осужденные за убийство женщины чаще всего защищались от **домашнего насилия**

Доля дел, в которых обвиняемые обронялись от партнёра

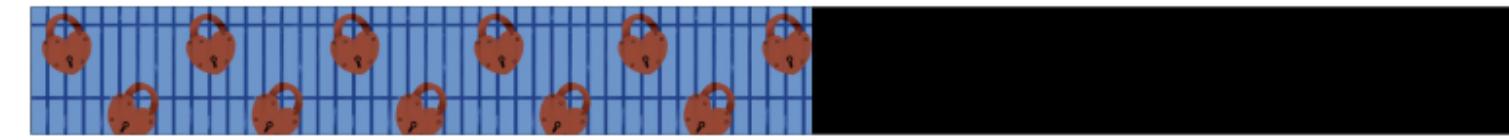
ч. 1 ст. 105 УК

79%



ч. 4 ст. 111 УК

52%



Данные: анализ приговоров женщинам по части 1 статьи 105 УК (убийство) и
части 4 статьи 111 УК (умышленное причинение тяжкого вреда здоровью,
повлекшее смерть), вынесенных в 2016-2018 годах; ГАС РФ «Правосудие».

<https://novayagazeta.ru/articles/2019/11/25/82847-ya-tebya-seychas-suka-ubivat-budu>

АЛГОРИТМЫ КЛАССИФИКАЦИИ В ЖУРНАЛИСТИКЕ



Журналисты с
помощью
алгоритма смогли
вычислить
«шпионские»
самолёты

<https://www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes>

Данные: сайт Flightradar24 — 20 000 самолётов

Скорость поворота,
Высота полёта,
Производитель,
Модель самолёта,
Траектория полёта,
Длительность полётов



сведения о том,
является ли самолёт
шпионским:

часть из них
известны, часть —
выясняет алгоритм

признаки или фичи (Features)

ПРЕДСКАЗАНИЕ ЗНАЧЕНИЙ (ИЛИ РЕГРЕССИЯ)



РЕГРЕССИЯ В ЖУРНАЛИСТИКЕ



Проект, который показал, что киноакадемия действительно часто присуждает «Оскар» по формальным причинам

<https://ria.ru/20200207/1564342281.html>

Данные: imdb.com, 12 тысяч фильмов

Жанр

Ключевые слова

Дата выхода

Предыдущие номинации
актёров, режиссёров,
сценаристов

признаки или фичи (Features)



количество
номинаций на
«Оскаре»

(алгоритм обучился на
небольшой группе
фильмов, чтобы
посчитать оценку для
каждого)

НУЖНО ЛИ МНЕ МАШИННОЕ ОБУЧЕНИЕ?

НЕТ

- задачу можно решить алгоритмическим способом
- человек быстрее и проще выполнит ту же самую работу

ДА

- данных так много, что их трудно обработать вручную
- нет очевидного алгоритма для решения задачи

Если интересно больше:

https://vas3k.ru/blog/machine_learning/