

# ПРЕДИСЛОВИЕ

В этой книге я представил несколько репрезентативных методов машинного обучения и попытался по возможности просто изложить их суть. Ее предполагаемая аудитория – те, кто только начинает знакомство с машинным обучением и уже владеет математикой на уровне первых курсов университета. Но если вы не дружите с математикой, то можете ознакомиться с разъяснениями в конце каждой главы и примерно понять, какие задачи решаются с помощью этих методов.

Особенность данной книги в том, что в начале каждой главы ставится задача, а затем постепенно объясняются методы машинного обучения, необходимые для ее решения. В таблице ниже перечислены задачи и методы, которые будут рассматриваться в каждой главе.

Глава	Задача	Метод
1	Прогноз количества участников мероприятия	Линейная регрессия
2	Определение вероятности заболевания диабетом	Логистическая регрессия, решающее дерево
3	Оценка результатов обучения	Метод проверки на резервированных данных, перекрестная проверка
4	Сортировка винограда	Сверточная нейронная сеть
5	Определение вероятности заболевания диабетом (повтор)	Ансамблевые методы
6	Рекомендация события	Кластерный анализ, матричное разложение

В каждой главе будет предложено лишь введение в тот или иной метод. Если вы хотите применить его на практике для решения какой-либо задачи, я советую обратиться к учебникам, которые указаны в списке рекомендованной литературы в конце книги.

Наконец, я благодарю всех сотрудников издательства Ohmsha за возможность написать эту книгу. Я также благодарен г-же Ватари Макана и всем сотрудникам Уильтэ, которые превратили мою рукопись в веселую мангу.

Июль 2018 года,  
*Араки Масахиро*

# СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ .....	V
-------------------	---

## Пролог

ПОГОВОРИМ О МАШИННОМ ОБУЧЕНИИ .....	1
-------------------------------------	---

В кабинете у Саяка (1). Саяка и старшеклассница Ай .....	14
--	----

## Глава 1

ЧТО ТАКОЕ РЕГРЕССИЯ .....	15
---------------------------	----

1.1. Сложности с прогнозом .....	16
----------------------------------	----

1.2. Определяем зависимые и независимые переменные .....	17
--	----

1.3. Находим функцию линейной регрессии .....	20
---	----

1.4. Регуляризация результата .....	22
-------------------------------------	----

В кабинете у Саяка (2). Математическое повторение (1) .....	34
---	----

## Глава 2

КАК ДЕЛАТЬ КЛАССИФИКАЦИЮ? .....	39
---------------------------------	----

2.1. Приводим данные в порядок .....	46
--------------------------------------	----

2.2. Определяем класс данных .....	47
------------------------------------	----

2.3. Логистическая регрессия .....	49
------------------------------------	----

2.4. Классификация по решающему дереву .....	55
--	----

В кабинете у Саяка (3). Математическое повторение (2) .....	74
---	----

## **Глава 3**

### **ОЦЕНКА РЕЗУЛЬТАТОВ ..... 77**

3.1. Без проверки тестовых данных никак нельзя.....	82
3.2. Обучающая, тестовая и проверочная выборки .....	83
3.3. Метод перекрестной проверки (кросс-валидации) .....	85
3.4. Доля правильно предсказанных объектов, точность, полнота и F-мера.....	87
В кабинете у Саяка (4). Математическое повторение (3).....	95

## **Глава 4**

### **ГЛУБОКОЕ ОБУЧЕНИЕ ..... 97**

4.1. Нейронная сеть .....	103
4.2. Обучение методом обратного распространения ошибок .....	107
4.3. Вызовы глубокого обучения.....	111
4.3.1. Проблема глубокой нейронной сети .....	112
4.3.2. Хитрости многоступенчатого обучения	
1. Метод предварительного обучения .....	113
4.3.3. Хитрости многоступенчатого обучения	
2. Функция активации.....	115
4.3.4. Хитрости многоступенчатого обучения	
3. Как избежать переобучения.....	117
4.3.5. Нейронные сети со специализированной структурой.....	118
В кабинете у Саяка (5). Математическое повторение (4) .....	134

## **Глава 5**

### **АНСАМБЛЕВЫЕ МЕТОДЫ ..... 139**

5.1. Бэггинг .....	146
5.2. Случайный лес.....	149
5.3. Бустинг .....	152
В кабинете у Саяка (6). Математическое повторение (5) .....	160

<b>Глава 6</b>	
<b>ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ .....</b>	<b>165</b>
6.1. Кластеризация.....	172
6.1.1. Иерархическая кластеризация .....	173
6.1.2. Разделяющая кластеризация.....	175
6.2. Разложение матрицы.....	179
В кабинете у Саяка (7). Математическое повторение (6) .....	191
<b>ЭПИЛОГ .....</b>	<b>197</b>
<b>ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ.....</b>	<b>205</b>

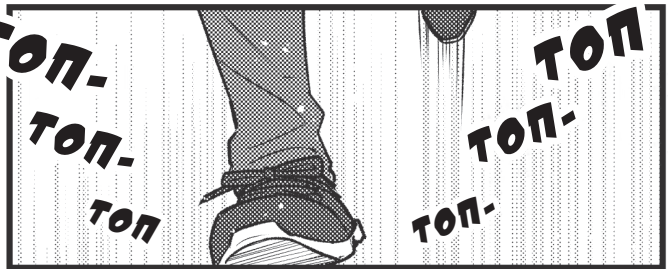
ПРОЛОГ

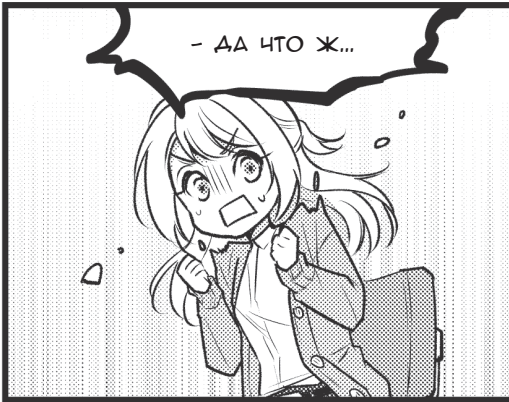
# ПОГОВОРИМ О МАШИННОМ ОБУЧЕНИИ

ЗАЧЕМ НУЖНО  
МАШИННОЕ ОБУЧЕНИЕ?



В одном университете





- ДА ЧТО Ж...



...ТАКОЕ! КИЁХАРА-КУН,  
ЧТО СЛУЧИЛОСЬ?  
ДАВНО ТЕБЯ  
НЕ ВИДЕЛА!

- А, СЯКА-  
СЭМПАЙ!

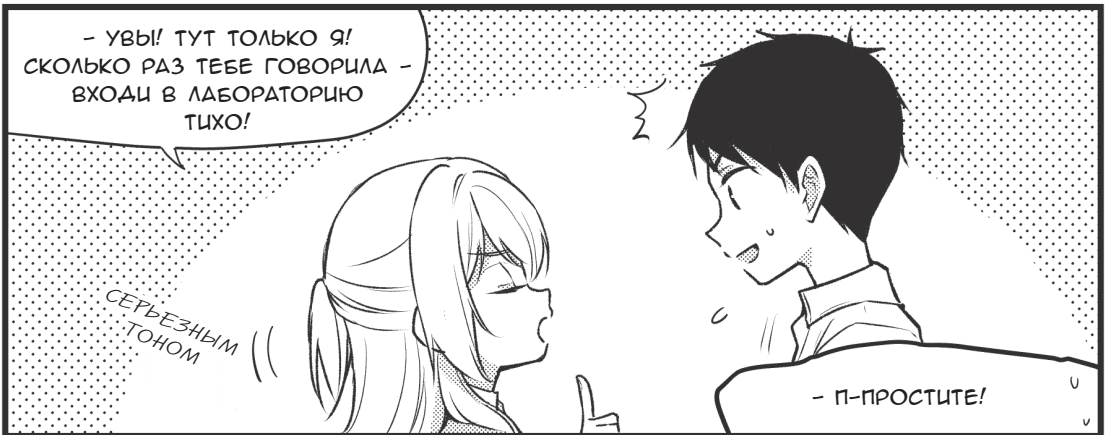


- СЭМПАЙ, ТУТ ЕСТЬ КТО-НИБУДЬ ЕЩЕ?  
НАДО ПОВОРОТИТЬ  
С ПРОФЕССОРОМ НАМИГОЗ...

**Киэхара Кадзума**

Вот уже год как работает в городской администрации.

Изучал машинное обучение в университете на факультете компьютерных технологий, но так ничему толком не выучился.



- УВЫ! ТУТ ТОЛЬКО Я!  
СКОЛЬКО РАЗ ТЕБЕ ГОВОРИЛА -  
ВХОДИ В ЛАБОРАТОРИЮ  
ТИХО!

СЕРЬЕЗНЫМ  
ТОНОМ

- П-ПРОСТИТЕ!

Сэмпай – название старших по курсу/классу учеников или студентов. Антоним – слово «кохай», которое обозначает младшего по курсу. – Прим. перев.

КИЁХАРА-КУН СОВСЕМ  
НЕ ПОМЕНЯЛСЯ,  
ХОТЬ ТЕПЕРЬ  
И НАШЕЛ РАБОТУ...

ВСЕ ТАКОУ ЖЕ  
ШУМНЫИ, БЕСТОЛКО-  
ВЫИ, НЕ СЛЫШИШЬ,  
ЧТО ТЕБЕ ГОВОРЯТ,  
И ПОТОМ...

**Мияно Саяка**  
Сэмпай Киёхары.  
Учится на втором курсе  
магистратуры.

САЯКА-СЭМПАИ,  
ВЫ КАК ОБЫЧНО...

ЧТО КАК ОБЫЧНО,  
КАК ОБЫЧНО  
ПРИДИРАЮСЬ...

НЕТ, НЕТ!

КАК ОБЫЧНО,  
МИЛАЯ...

НЕТ, ЭТО НЕ ТАК.  
А КОГДА ВЕРНЕТСЯ  
ПРОФЕССОР НАМИГОЭ?

ВОТ  
ПОДАРОК

ОН ВМЕСТЕ С ДРУГИМИ  
ЛАБОРАНТАМИ В КОМАНДИРОВКЕ  
ЗА ГРАНИЦЕЙ, РАНЬШЕ, ЧЕМ  
ЧЕРЕЗ ДВА МЕСЯЦА,  
НЕ ВЕРНЕТСЯ.

СПАСИБО!

ЧТО?!

Угу  
Ну мне  
так кажется...





БЛИН, МНЕ ЭТО НИКАК НЕ ПОДХОДИТ!

ЧТО-ТО СЛУЧИЛОСЬ?

ТРЯСЕТСЯ

В УЖАСЕ



Я ХОТЕЛ ПОПРОСИТЬ ПРОФЕССОРА НАМИГОЭ РАССКАЗАТЬ МНЕ ПРО МАШИННОЕ ОБУЧЕНИЕ...

МАШИННОЕ ОБУЧЕНИЕ?



КИЁХАРА, ТЫ ЖЕ РАБОТАЕШЬ У СЕБЯ В МЕСТНОЙ АДМИНИСТРАЦИИ. ТЫ ЖЕ ГОВОРИЛ: "ХОЧУ СПОКОЙНО РАБОТАТЬ ДОМА И ЖИТЬ В СВОЕ УДОВОЛЬСТВИЕ". ЗАЧЕМ ТЕБЕ МАШИННОЕ ОБУЧЕНИЕ?

Киёхара-кун в университете

О-ХО-ХО...

Что? Работу искать? Хочу не слишком напряженно работать в хорошем месте



У ТЕБЯ ЧТО-ТО СТЯСЛОСЬ? МОЖЕТ, МНЕ РАССКАЖЕШЬ?

.....

У НАС ЕСТЬ  
КОНСУЛЬТАНТ, КОТОРЫЙ ПОРУЧИЛ  
МНЕ ПОРАБОТАТЬ С ЕГО ЦИ-ПРОГРАМ-  
МОЙ, ПРОГНОЗИРУЮЩЕЙ КОЛИЧЕСТВО  
ГОСТЕЙ НА ПУБЛИЧНЫХ МЕРОПРИЯТИЯХ  
НАШЕЙ АДМИНИСТРАЦИИ.

ЭТОТ КОНСУЛЬТАНТ -

К МОМЕНТУ НАСТУПЛЕНИЯ  
СИНГУЛЯРНОСТИ ПОЯВИТСЯ  
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МОЗГ,  
СРАВНИМЫЙ С ЧЕЛОВЕЧЕСКИМ  
ИЛИ ДАЖЕ ПРЕВОСХОДЯЩИЙ ЕГО,  
КОТОРЫЙ СДЕЛАЕТ ВСЮ ЧЕЛОВЕ-  
ЧЕСКУЮ РАБОТУ НЕНУЖНОЙ!

**ВОДУШЕВЛЕННО**

...ПРИМЕРНО ТАКОЙ ТИП...  
Я ПОДСЧИТАЛ КОЛИЧЕСТВО ГОСТЕЙ  
ЧЕРЕЗ ЦИ-ПРОГРАММУ, КОТОРУЮ  
ОН РАЗРАБОТАЛ...



СУДЯ ПО ПРОГНОЗАМ ЦИ,  
КОЛИЧЕСТВО ГОСТЕЙ  
УМЕНЬШИТСЯ!

СПАСИБО!



ЦИФРЫ БЫЛИ ПОДОЗРИ-  
ТЕЛЬНЫМИ, И КОГДА  
Я ЦИ ПРОВЕРИЛ...



...ТО УВИДЕЛ,  
ЧТО КОЛИЧЕСТВО ГОСТЕЙ  
ДВА ПОСЛЕДНИХ ГОДА  
СНИЖАЕТСЯ ПО ПРЯМОЙ!

Я ДОЛОЖИЛ ОБ ЭТОМ  
ОТВЕТСТВЕННОМУ  
ЗА РЕКЛАМУ, НО ТОТ  
НЕ ОБРАТИЛ  
ВНИМАНИЯ...



ПОЭТОМУ Я ПОДУМАЛ, ЧТО  
ЕСЛИ БЫ СМОГ, ИСПОЛЬЗУЯ  
ДААННЫЕ ЗА ДЕСЯТЬ ЛЕТ,  
СДЕЛАТЬ ПРОГНОЗ  
ПРИ ПОМОЩИ МАШИННОГО

ОБУЧЕНИЯ, ТО У МЕНЯ  
ПОЛУЧИЛОСЬ БЫ ЕГО  
УБЕДИТЬ...

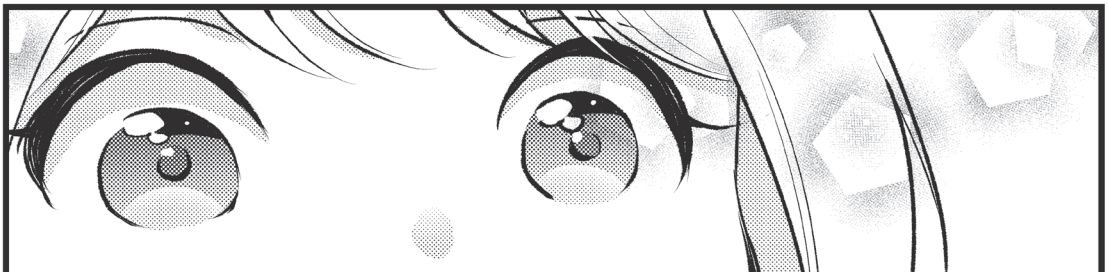
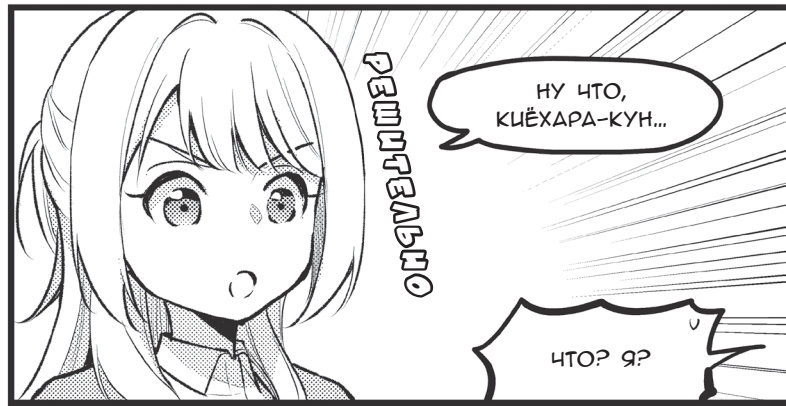
МОЖЕТ,  
НА ЭТО КОЛИЧЕСТВО  
ОСАДКОВ В СЕЗОН  
ДОЖДЕЙ ВЛИЯЕТ?..

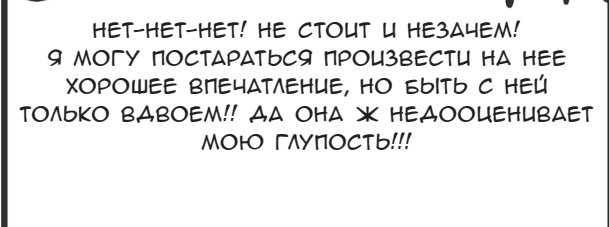


НО Я  
НИЧЕГО УМЕЮ...

ПОЭТОМУ Я ПРИШЕЛ  
К ПРОФЕССОРУ  
НАМИГОЗ, ЧТОБЫ ОН  
РАССКАЗАЛ МНЕ  
О МАШИННОМ  
ОБУЧЕНИИ...

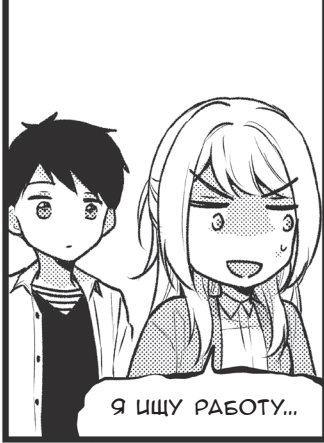








КСТАТИ, СЯКА-СЭМПАЙ,  
А ПОЧЕМУ ВЫ  
НЕ В КОМАНДИРОВКЕ  
С УЧИТЕЛЕМ НАМИГОЭ?



Я ИЩУ РАБОТУ...



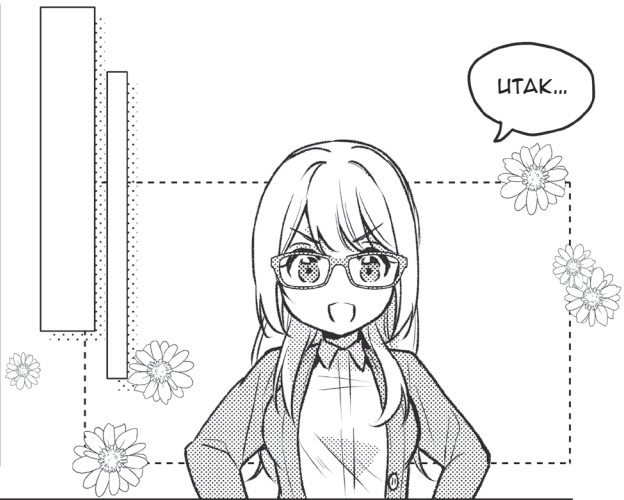
А РАЗВЕ ВЫ  
НЕ ДОЛЖНЫ ХОДИТЬ  
НА СОБЕСЕДОВАНИЯ?

ПОМОЛЧИ-КА!



Я ПРОСТО НЕ МОГУ ПРЕПОДАВАТЬ  
ТЕМ, КТО ГОВОРИТ ТАКИЕ ВЕЩИ!

ПРОСТИТЕ! БУДЬТЕ  
СНИСХОДИТЕЛЬНЫ!



ИТАК...



ПРЕЖДЕ ВСЕГО, КИЁХАРА-КУН,  
ДАВАЙ ПРОВЕРИМ, ЧТО ТЫ ЗНАЕШЬ  
О МАШИННОМ ОБУЧЕНИИ.



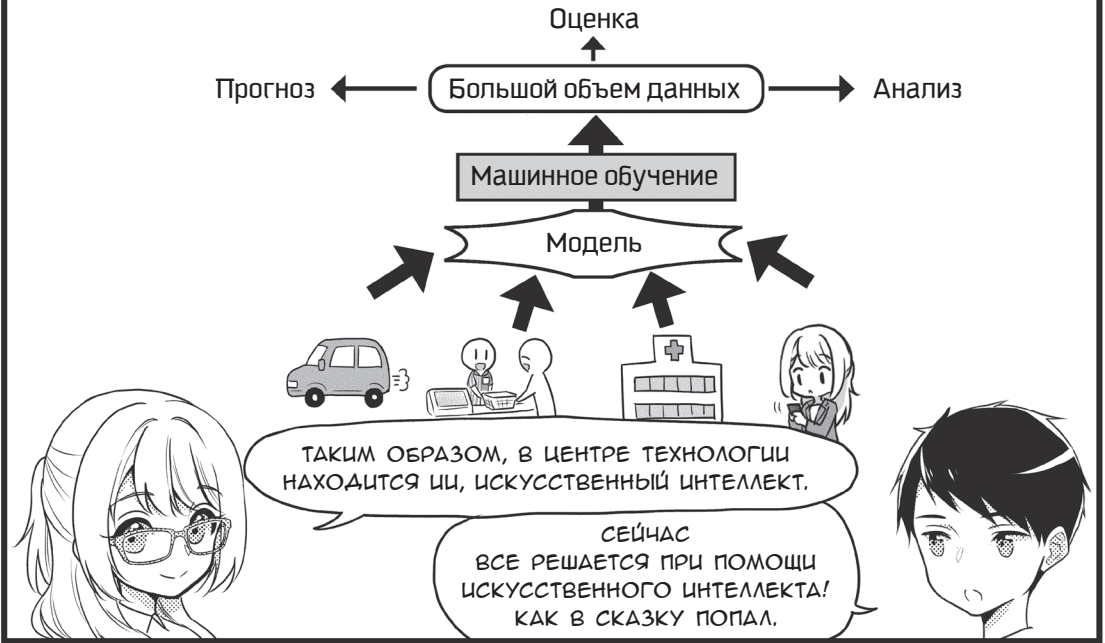
НУ... ЭТО КОГДА ОН АНАЛИЗИРУЕТ  
БОЛЬШОЙ ОБЪЕМ ДАННЫХ  
И ДАЕТ ОТВЕТ?



ОЧКИ...

НУ ЧТО, КИЁХАРА...  
ИТАК, МАШИННОЕ ОБУЧЕНИЕ - ЭТО

ПОСТРОЕНИЕ НА ОСНОВАНИИ БОЛЬШОГО ОБЪЕМА ДАННЫХ  
МОДЕЛИ, КОТОРАЯ МОЖЕТ ОЦЕНИВАТЬ И ДЕЛАТЬ ПРОГНОЗЫ.



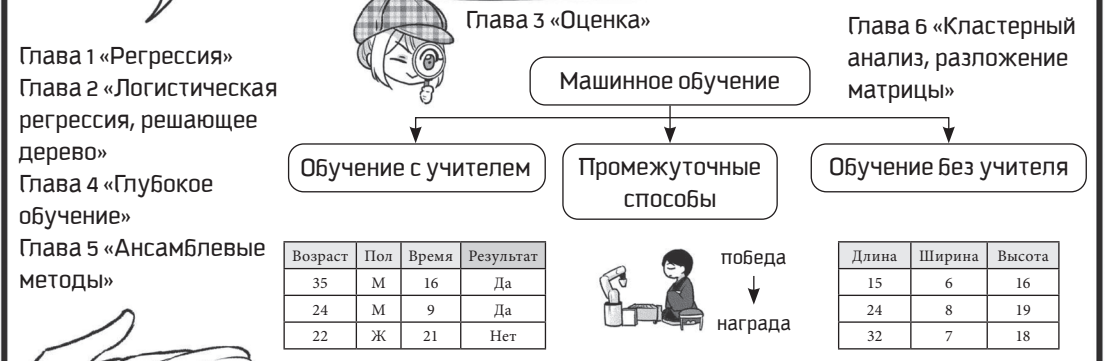
ИИ СЕЙЧАС ДОСТАТОЧНО РАСПРОСТРАНЕН! В ЦЕЛОМ СЧИТАЕТСЯ, ЧТО ОН ВСКОРЕ ЗАМЕНИТ ЛЮДЕЙ ПРИ ВЫПОЛНЕНИИ НЕКОТОРЫХ ПРОСТЫХ ИНТЕЛЛЕКТУАЛЬНЫХ ЗАДАНИЙ, НО В ДРУГИХ СИТУАЦИЯХ ОН ПОМОЖЕТ РАСШИРИТЬ ВОЗМОЖНОСТИ ЧЕЛОВЕЧЕСКОГО УМА.

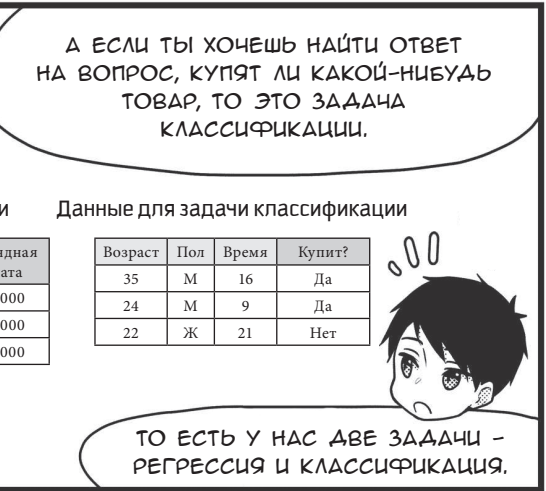
ОГО.

БОЛЕЕ ТОГО, "МАШИННОЕ ОБУЧЕНИЕ" ТЕСНО СВЯЗАНО С ТЕХНОЛОГИЕЙ DATA MINING (ДОБЫЧА ДАННЫХ), КОТОРАЯ ПОЗВОЛЯЕТ ПОЛУЧИТЬ НУЖНЫЕ ДАННЫЕ В РЕЗУЛЬТАТЕ АНАЛИЗА ОГРОМНОГО, НЕПРЕДСТАВИМОГО ДЛЯ ЧЕЛОВЕЧЕСКОГО РАЗУМА ОБЪЕМА ДАННЫХ.

**Data mining** - метод обнаружения скрытых паттернов в огромном объеме данных при помощи статистики и математических методов.

ПОСКОЛЬКУ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ РЕШАЕТСЯ ОГРОМНОЕ КОЛИЧЕСТВО ЗАДАЧ, ЕГО ОБОЗРЕТЬ В ЦЕЛОМ ТРУДНО, ОДНАКО МЕТОДЫ ЛЕГКО РАЗДЕЛИТЬ ПРИМЕРНО НА ТРИ ГРУППЫ: **ОБУЧЕНИЕ С УЧИТЕЛЕМ; ПРОМЕЖУТОЧНЫЕ СПОСОБЫ; ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ.**





Данные для задачи регрессии

Кол-во комнат	Время ходьбы от станции, мин	Возраст дома	Арендная плата
1	15	6	48 000
2	2	2	60 000
3	20	25	50 000

Данные для задачи классификации

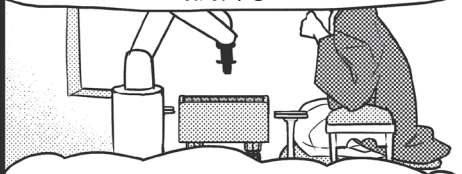
Возраст	Пол	Время	Купит?
35	М	16	Да
24	М	9	Да
22	Ж	21	Нет

ТО ЕСТЬ У НАС ДВЕ ЗАДАЧИ - РЕГРЕССИЯ И КЛАССИФИКАЦИЯ.



ПРОЛОГ. ПОВОРОМ О МАШИННОМ ОБУЧЕНИИ

ЗАТЕМ ИДУТ ПРОМЕЖУТОЧНЫЕ МЕТОДЫ. КИЁХАРА-КУН, ТЫ СЛЫШАЛ КОГДА-НИБУДАЬ НОВОСТИ О ТОМ, ЧТО ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ОБЫГРАЛ ЧЕЛОВЕКА В ШАХМАТЫ ИЛИ ГО?



АГА, ЭТО БЫЛО ВОСХИТИТЕЛЬНО. ХОТЯ РАНЬШЕ КАЗАЛОСЬ, ЧТО ИИ ЭТОГО НЕ СМОЖЕТ.

МЕТОД, ИСПОЛЬЗУЕМЫЙ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ ДЛЯ ИГРЫ В ГО ИЛИ ШАХМАТЫ, НАЗЫВАЕТСЯ ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ.



車	馬	龍	龍	王	龍	龍	馬	車
	車						車	
歩	歩	歩	歩	歩	歩	歩	歩	歩
	角						飛	
龍	馬	銀	金	玉	金	銀	馬	龍

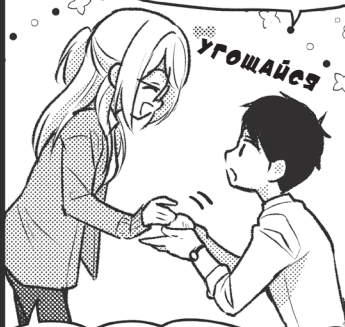
Победа – положительное подкрепление

Проигрыш – отрицательное подкрепление

После каждого хода неизвестно, какой следующий ход лучше сделать\*

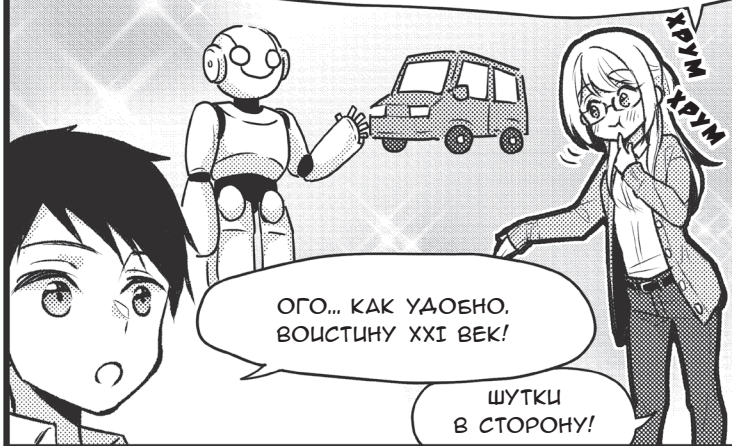
ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ НЕ ДАЕТ ОТВЕТА НА ВОПРОС, КАКОЙ ХОД ЛУЧШЕ СДЕЛАТЬ, НО ВМЕСТО ЭТОГО ВЫДАЕТ ПОДКРЕПЛЕНИЕ В ЗАВИСИМОСТИ ОТ ПОБЕДЫ ИЛИ ПОРАЖЕНИЯ.

ЛЮДЯМ ТОЖЕ МОЖНО ДАВАТЬ ПОДКРЕПЛЕНИЕ.



АА, ЭТИ СЛАДОСТИ – КАК РАЗ ПОДКРЕПЛЕНИЕ ДЛЯ МЕНЯ.

НА ОСНОВАНИИ ПОДКРЕПЛЕНИЯ ОПРЕДЕЛЯЕТСЯ ОПТИМАЛЬНЫЙ ПОРЯДОК ДЕЙСТВИЙ. ТАК ОБУЧАЮТ РОБОТОВ ДЛЯ ВОЖДЕНИЯ АВТОМОБИЛЕЙ.



ОГО... КАК УДОБНО. ВОИСТИНУ XXI ВЕК!

ШУТКИ В СТОРОНУ!

\* Приведенная таблица – расстановка для игры в сёги, японские шахматы. – Прим. перев.



ЧТО ЖЕ КАСАЕТСЯ ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ, ТО ЖЕЛАЕМОГО ОТВЕТА НА ВОПРОСЫ СРЕДИ ДАННЫХ НЕТ.

Обучение без учителя

Неразмеченные данные

Длина	Ширина	Высота
15	6	16
24	8	19
32	7	18

А КАК ТОГДА ПРОХОДИТ ОБУЧЕНИЕ?

НУ, КАК ОБЫЧНО. ЦЕЛЬ СИСТЕМ БЕЗ УЧИТЕЛЯ - ЭТО ОБНАРУЖИТЬ В БОЛЬШОМ ОБЪЕМЕ ДАННЫХ ЗНАНИЯ, КОТОРЫЕ МОГУТ ПРИГОДИТЬСЯ ЧЕЛОВЕКУ.

ОНИ ИСПОЛЗУЮТСЯ, НАПРИМЕР, ДЛЯ РЕКОМЕНДАЦИИ ТОВАРОВ ПРИ ПОКУПКЕ В ИНТЕРНЕТ-МАГАЗИНАХ ИЛИ ПРИ ПОИСКЕ СТРАННОСТЕЙ В ИНФОРМАЦИИ О ДЕЙСТВИЯХ МЕХАНИЗМОВ.

История покупок

ID	#1	#2	#3	#4
115		1		
124		1		1
232				1

Рекомендуемые артисты



ПОКА ЧТО, НАДЕЮСЬ, ТЫ ВСЕ ПОНЯЛ ОТНОСИТЕЛЬНО МАШИННОГО ОБУЧЕНИЯ?

ДА...

ЭТО УЖЕ ВОСЬМОЙ.

## В кабинете у Саяка (1)

### Саяка и старшеклассница Ай



Давно не виделись, Ай-тян. С тех пор как мы были у дедушки, да?

Возможно, тогда собирались все его внуки. Чем ты сегодня занималась?



Рассказывала младшему товарищу про машинное обучение. Он хоть и учился у нас, но, боюсь, очень мало что понял... Ай-тян, ты же хочешь учиться в математическом классе, может, тоже зайдешь послушать?

Машинное обучение? ИИ? Это, наверное, очень сложно!



Там используются математические модели, но это все делается компьютерами. А основы этих моделей может понять и старшеклассник.

Я писала программы для астрономических расчетов. Математика – мой любимый предмет, но пойму ли я?



Конечно, поймешь. Мы начнем с регрессии. Приходи послушать!

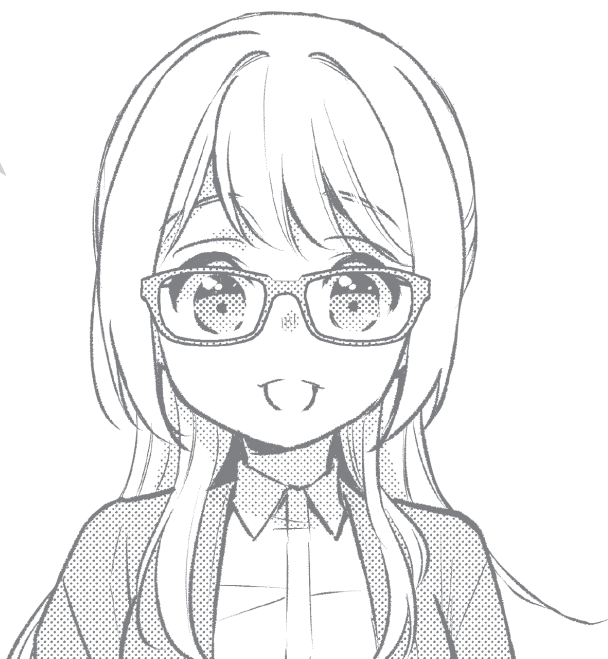
Хорошо, если что-то будет трудно, я буду задавать вопросы!



## ГЛАВА 1

# ЧТО ТАКОЕ РЕГРЕССИЯ

ЛИНЕЙНАЯ РЕГРЕССИЯ!  
РЕГУЛЯРИЗАЦИЯ!



Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

[e-Univers.ru](http://e-Univers.ru)