

УДК 004.4
ББК 32.972
А79

Араки М., Ватари М.

А79 Занимательная манга. Машинное обучение: манга / Араки Масахиро (автор), Ватари Макана (худ.); пер. с яп. А. С. Слащевой. — М.: ДМК Пресс, 2020. — 214 с. : ил. — (Серия «Образовательная манга»). — Доп. тит. л. яп.

ISBN 978-5-97060-830-2

Сотруднику городской администрации Киёхара Кадзума поручено задание, которое без машинного обучения не выполнить. Под руководством своей давней знакомой Мияно Саяка он осваивает премудрости работы с искусственным интеллектом – от самых азов до глубокого обучения.

Вместе с героями манги читатели узнают о том, что такое регрессия и как проводить классификацию, ознакомятся с принципами оценки тестовых данных и особенностями работы нейронных сетей. В заключительной части излагаются методы обучения без учителя.

Издание предназначено для тех, кто начинает знакомство с машинным обучением и освоил математику на уровне первых курсов университета.

УДК 004.4
ББК 32.972

Manga de Wakaru Kikai Gakushu (Manga Guide: Machine Learning)

By Araki Masaxiro (Author), Illustration by Vatarai Makana

Published by Ohmsha, Ltd.

Russian language edition copyright © 2020 by DMK Press

Все права защищены. Никакая часть этого издания не может быть воспроизведена в любой форме или любыми средствами, электронными или механическими, включая фотографирование, ксерокопирование или иные средства копирования или сохранения информации, без письменного разрешения издательства.

ISBN 978-4-274-22244-3 (яп.)

ISBN 978-5-97060-830-2 (рус.)

Copyright © 2018 by and Office sawa, Ltd.

© Издание, перевод, ДМК Пресс, 2019

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

Глава 6	
ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ	165
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
ЭПИЛОГ	197
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ.....	205

ПРОЛОГ

ПОГОВОРИМ О МАШИННОМ ОБУЧЕНИИ

ЗАЧЕМ НУЖНО
МАШИННОЕ ОБУЧЕНИЕ?



В одном университете






Киёхара Кадзума

Вот уже год как работает в городской администрации.


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



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
ШУМНЫУ, БЕСТОЛКО-
ВУУ, НЕ СЛЫШИШЬ,
ЧТО ТЕБЕ ГОВОРЯТ,
И ПОТОМ...

Мияно Саяка
Сэмпей Киёхары.
Учится на втором курсе
магистратуры.


САЯКА-СЭМПАУ,
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.

Угу
Ну мне
так кажется...



СПАСИБО!

ЧТО?!



БЛИН, МНЕ ЭТО
НИКАК НЕ ПОДХОДИТ!

ЧТО-ТО
СЛУЧИЛОСЬ?

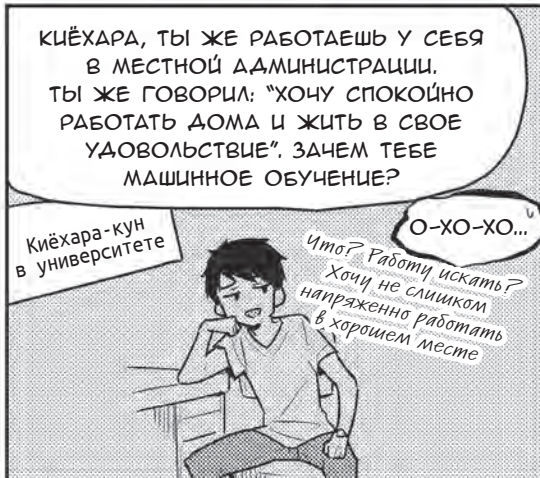
ТРЯСЕТСЯ

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!

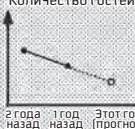
ВОДУШЕВЛЕННО

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

СУДЯ ПО ПРОГНОЗАМ ЦИ,
КОЛИЧЕСТВО ГОСТЕЙ
УМЕНЬШИТСЯ!

СПАСИБО!

Количество гостей



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

ой!

...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...







НЕТ-НЕТ, СЭМПАИ, ВЫ ЖЕ ТАК ЗАНЯТЫ!

Я И САМ ЧТО-НИБУДЬ ПРИДУМАЮ, А-А-А-А-А!

НЕТ-НЕТ-НЕТ! НЕ СТОИТ И НЕЗАЧЕМ!
Я МОГУ ПОСТАРАТЬСЯ ПРОИЗВЕСТИ НА НЕЕ ХОРОШЕЕ ВПЕЧАТЛЕНИЕ, НО БЫТЬ С НЕЙ ТОЛЬКО ВАВОЕМ!! ДА ОНА Ж НЕДООЦЕНИВАЕТ МОЮ ГЛУПОСТЬ!!!





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



Я ИЩУ РАБОТУ...



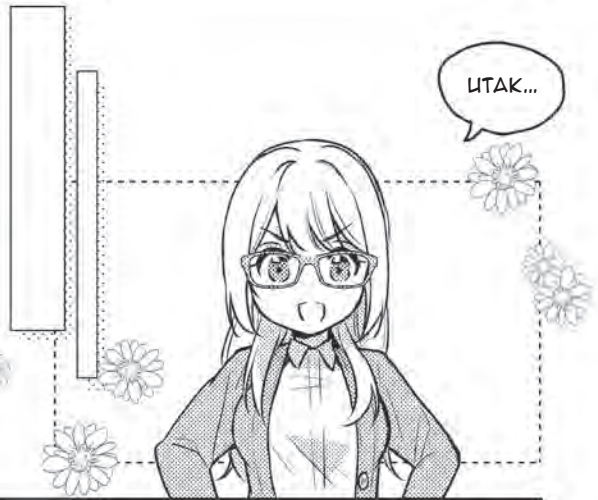
А РАЗВЕ ВЫ
НЕ ДОЛЖНЫ ХОДИТЬ
НА СОБЕСЕДОВАНИЯ?

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.



Машинное обучение

Модель



ТАКИМ ОБРАЗОМ, В ЦЕНТРЕ ТЕХНОЛОГИИ НАХОДИТСЯ ИИ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ.

СЕЙЧАС
ВСЕ РЕШАЕТСЯ ПРИ ПОМОЩИ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА!
КАК В СКАЗКУ ПОПАЛ.

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.**

- Глава 1 «Регрессия»
- Глава 2 «Логистическая регрессия, решающее дерево»
- Глава 4 «Глубокое обучение»
- Глава 5 «Ансамблевые методы»



Глава 3 «Оценка»

Машинное обучение

Глава 6 «Кластерный анализ, разложение матрицы»

Обучение с учителем

Промежуточные способы

Обучение без учителя

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



победа
↓
награда

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





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

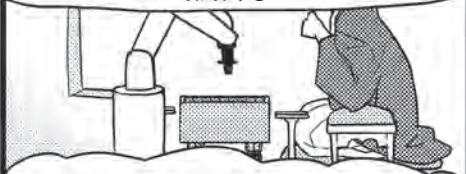
Возраст	Пол	Время	Купит?
35	М	16	Да
24	М	9	Да
22	Ж	21	Нет

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 НИ ЭТОГО НЕ СМОЖЕТ.

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



香車	香車	桂馬	桂馬	桂馬	桂馬	桂馬	桂馬	桂馬	香車
	香車							桂馬	
香車	香車	香車	香車	香車	香車	香車	香車	香車	香車
歩兵	歩兵	歩兵	歩兵	歩兵	歩兵	歩兵	歩兵	歩兵	歩兵
								龍馬	
香車	桂馬	銀将	金将	玉将	金将	銀将	桂馬	香車	

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.



* 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!



ДЛЯ НАЧАЛА
ПОГОВОРИМ
О РЕГРЕССИИ.

ПОЖАЛУЙСТА!

УВАЖАЮ

В КАЧЕСТВЕ ПРИМЕРА
ВОЗЬМЕМ СИТУАЦИЮ, КОГДА НАДО
ПОДСЧИТАТЬ КОЛИЧЕСТВО ГОСТЕЙ
НА ПИАР-МЕРОПРИЯТИИ, КОТОРОЕ
ОРГАНИЗОВЫВАЕТ ГОРОД.

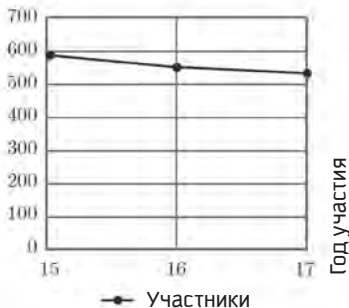
НА НЕМ БУДЕТ ПОДАВАТЬСЯ СОК,
ВЫЖАТЫЙ ИЗ МЕСТНЫХ ФРУКТОВ,
ПОЭТОМУ НАДО КАК МОЖНО ТОЧНЕЕ
ПОДСЧИТАТЬ КОЛИЧЕСТВО
УЧАСТНИКОВ.

АГА!

1.1. СЛОЖНОСТИ С ПРОГНОЗОМ

ДЛЯ НАЧАЛА ПОСМОТРИМ
НА ЭТОТ ГРАФИК.

Количество участников

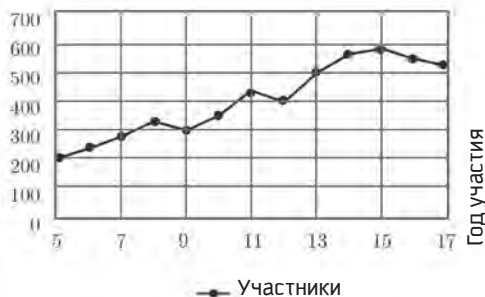


ЭТОТ ГРАФИК ПОКАЗЫВАЕТ
ИЗМЕНЕНИЕ ЧИСЛА УЧАСТНИКОВ
ЗА ТРИ ГОДА. ЕСЛИ МЫ
ПРЕДСКАЖЕМ КОЛИЧЕСТВО
УЧАСТНИКОВ НА ЕГО ОСНОВЕ,
ЧТО ПОЛУЧИТСЯ?

ЗА ТРИ ГОДА
ДЕМОНСТРИРУЕТСЯ
ТЕНДЕНЦИЯ К СНИЖЕНИЮ.

ДА, А ЕСЛИ МЫ ВЗГЛЯДЕМ
НА ГРАФИК УЧАСТНИКОВ
С 13-ГО ГОДА?

Количество участников



В ДОЛГОСРОЧНОЙ
ПЕРСПЕКТИВЕ ВИДНА
ТЕНДЕНЦИЯ К УВЕЛИЧЕНИЮ
ЧИСЛА УЧАСТНИКОВ.

ИМЕННО ТАК! В ЗАВИСИМОСТИ ОТ ТОГО, КАКОЙ ПЕРИОД МЫ ВЗЯЛИ, ТЕНДЕНЦИИ АБСОЛЮТНО ПРОТИВОПОЛОЖНЫЕ.



И ЭТО ТОТ СПОСОБ, КОТОРЫЙ ИСПОЛЬЗОВАЛ КОНСУЛЬТАНТ?

ДА.

ДАЖЕ ЕСЛИ ТЕНДЕНЦИЯ И ПРАВИЛЬНАЯ, ГРАФИК НЕОБЯЗАТЕЛЬНО БУДЕТ ПОХОДИТЬ НА ПРЯМУЮ.

ВООДУШЕВЛЕННО



1.2. ОПРЕДЕЛЯЕМ ЗАВИСИМЫЕ И НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ

ЧТО Ж, КИЁХАРА, КАК МЫ ТЕПЕРЬ БУДЕМ ПРЕДСКАЗЫВАТЬ КОЛИЧЕСТВО ГОСТЕЙ?

СТАРАТЕЛЬНО ЗАПИСЫВАЕТ



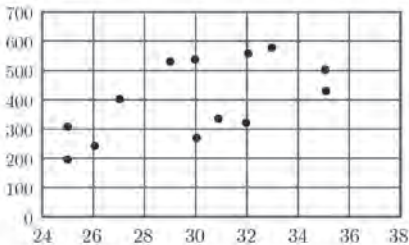
НУ... Я ДУМАЮ, ВЛИЯНИЕ ОКАЗЫВАЮТ ПОГОДА, ТЕМПЕРАТУРА, В ОСОБЕННОСТИ КОЛИЧЕСТВО ОСАДКОВ В СЕЗОН ДОЖДЕЙ...

А ТЕПЕРЬ ПОПРОБУЕМ ВЫБРАТЬ МЕТОД ПРЕДСКАЗАНИЯ В ЗАВИСИМОСТИ ОТ ФАКТОРОВ. ТЕ РЕЗУЛЬТАТЫ, КОТОРЫЕ МЫ ХОТИМ СПРОГНОЗИРОВАТЬ, НАЗЫВАЮТСЯ ЗАВИСИМЫМИ ПЕРЕМЕННЫМИ, А ФАКТОРЫ, ВЛИЯЮЩИЕ НА РЕЗУЛЬТАТ, - НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ.

ДА!



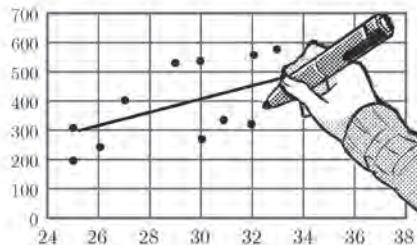
ПУСТЬ У НАС БУДЕТ ОДНА НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ. РАССМОТРИМ ВЛИЯНИЕ ТЕМПЕРАТУРЫ В ДЕНЬ МЕРОПРИЯТИЯ НА КОЛИЧЕСТВО ГОСТЕЙ.



НА ЭТОМ ГРАФИКЕ ГОРИЗОНТАЛЬНАЯ ОСЬ - НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ (ПОГОДА), А ВЕРТИКАЛЬНАЯ - КОЛИЧЕСТВО УЧАСТНИКОВ (ЗАВИСИМАЯ ПЕРЕМЕННАЯ). ОТМЕТИМ ТОЧКИ, РАВНЫЕ КОЛИЧЕСТВУ УЧАСТНИКОВ, И ПРОВЕДЕМ ПО НИМ ПРЯМУЮ.

ЭТО НЕВОЗМОЖНО?

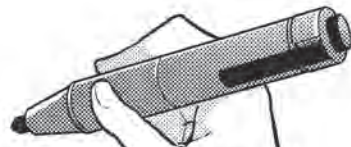
КАК-ТО ТАК. ПО ВОЗМОЖНОСТИ ПРОВОДИМ ЕЕ НЕДАЛЕКО ОТ ВСЕХ ТОЧЕК. КОГДА МЫ ПРОВОДИМ ЛИНИЮ, ОНА ВЫГЛЯДИТ ТАК:



СКРИП
СКРИП

ИТАК, ОБОЗНАЧИМ НАКЛОН ГРАФИКА w_1 ,
 СДВИГ ГРАФИКА ОТНОСИТЕЛЬНО
 ВЕРТИКАЛЬНОЙ ОСИ - w_0 , ТЕМПЕРАТУРУ ЗА x ,
 А КОЛИЧЕСТВО УЧАСТНИКОВ - ЗА y

$$y = w_1x + w_0$$



- И ПОЛУЧИМ ЭТУ ФОРМУЛУ.

ЕСЛИ ТЕМПЕРАТУРА РАСТЕТ,
 ТО И КОЛИЧЕСТВО УЧАСТНИКОВ
 РАСТЕТ, А ЭТО НЕПОХОЖЕ
 НА ПРАВДУ.



АГА.
 ВДИМО, НЕЗАВИСИМЫХ
 ПЕРЕМЕННЫХ НЕСКОЛЬКО.

НЕСКОЛЬКО?
 ЭТО УЖАСНО?

СЛОЖНОСТЬ В ТОМ, ЧТО СРЕДИ
 НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ ЕСТЬ ТАКИЕ,
 КОТОРЫЕ ОКАЗЫВАЮТ ВЛИЯНИЕ
 НА ЗАВИСИМЫЕ ПЕРЕМЕННЫЕ,
 А ЕСТЬ И ТАКИЕ, ВЛИЯНИЕ КОТОРЫХ
 НЕВЕЛИКО.

Влияние велико

Влияние мало

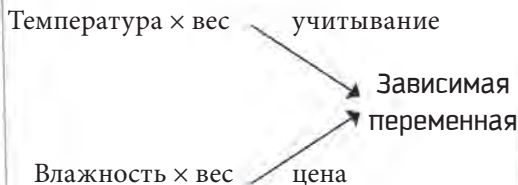
ПОГОДА

ВЛАЖНОСТЬ



ЧТОБЫ ПРОСТО СМОДЕЛИРОВАТЬ ТАКИЕ УСЛОВИЯ,
 НЕОБХОДИМ СПОСОБ, КОТОРЫЙ БУДЕТ УПОРЯДОЧИВАТЬ
 ЗАВИСИМЫЕ ПЕРЕМЕННЫЕ В ЗАВИСИМОСТИ ОТ УЧИТЫ-
 ВАЕМОГО ВЕСА НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ.

ЭТО
 ЛИНЕЙНАЯ РЕГРЕССИЯ.



Что такое линейная регрессия?



АГА! ВОТ И РЕГРЕССИЯ!

ЕСЛИ НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ ДВЕ,
ТО ОБОЗНАЧИМ ИХ КАК x_1 И x_2 ,
ИХ ВЕС КАК w_1 И w_2 СООТВЕТСТВЕННО,
А ИХ ВЗВЕШЕННАЯ СУММА ВМЕСТЕ
С ПОСТОЯННОЙ СОСТАВЛЯЮЩЕЙ w_0
ЗАПИСЫВАЕТСЯ...

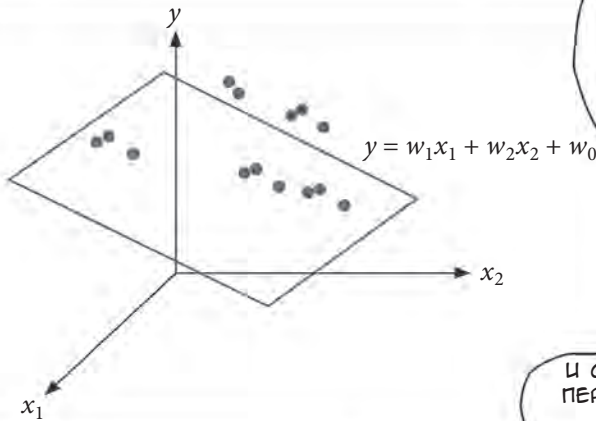
$$y = w_1x_1 + w_2x_2 + w_0$$

...ВОТ ТАК.

ОЧЕНЬ ПОХОЖЕ
НА НЕДАВНЮЮ ФОРМУЛУ.

СКРИП
СКРИП

МЫ ЖЕ СТРОИЛИ
АВУМЕРНЫЙ ГРАФИК?



МОЖНО ПОСТРОИТЬ ТРЕХМЕРНЫЙ
ГРАФИК, КАК НА КАРТИНКЕ,
И СПРОГНОЗИРОВАТЬ КОЛИЧЕСТВО
УЧАСТНИКОВ. ПОЛУЧИТСЯ ТАКАЯ
ВОТ ПОВЕРХНОСТЬ.

ФУХ!

И С КАЖДОЙ НОВОЙ НЕЗАВИСИМОЙ
ПЕРЕМЕННОЙ БУДЕТ УВЕЛИЧИВАТЬСЯ
КОЛИЧЕСТВО ИЗМЕРЕНИЙ?

ВСЕ ТАК... ОБОБЩАЕМ: ЕСЛИ
В ДАННЫХ СУЩЕСТВУЕТ
 d НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ,
ТО ТОЧКИ НА ГРАФИКЕ БУДУТ
РАСПОЛАГАТЬСЯ В $(d+1)$ -МЕРНОМ
ПРОСТРАНСТВЕ, И ПОЭТОМУ НАМ НАДО
РЕШИТЬ ЗАДАЧУ НАХОЖДЕНИЯ
 d -МЕРНОЙ ГИПЕРПЛОСКОСТИ.

ЧТО?

ТО ЕСТЬ ЕСЛИ НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ
БУДЕТ 10, ТО НАМ НУЖНО БУДЕТ НАЙТИ
ДЕСЯТИМЕРНУЮ ГИПЕРПЛОСКОСТЬ?
А ЭТО ВООООЩЕ ВОЗМОЖНО?

ДА! ПРИДЕТСЯ ПОПРОСИТЬ
ПОМОЩИ У МАТЕМАТИКИ!

1.3. НАХОДИМ ФУНКЦИЮ ЛИНЕЙНОЙ РЕГРЕССИИ

Шаг 1

Пусть у нас есть d независимых переменных. Тогда обозначим d -мерный столбец-вектор \mathbf{x} . В соответствии с этим вес, отличный от постоянной w_0 , также будет обозначен d -мерным вектором-столбцом \mathbf{w} , и уравнение гиперплоскости примет такой вид:

$$y = \mathbf{w}^T \mathbf{x} + w_0, \quad (1.1)$$

где T – обозначение транспонирования.

Шаг 2

До этого мы обозначили за y величину взвешенной суммы независимых переменных, а теперь обозначим так величину зависимых переменных в данных для обучения. Взвешенную сумму независимых переменных мы обозначим как $\hat{c}(\mathbf{x})$. Значок над буквой означает, что мы не можем гарантировать правильность полученных данных.

Таким образом уравнение (1.1) примет вид (1.2):

$$\hat{c}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0. \quad (1.2)$$



Тот факт, что функция линейной регрессии не слишком отличается от использованных данных, означает, что величина линейной функции $\hat{c}(\mathbf{x})$, куда входят независимые переменные \mathbf{x} , также мало отличается от величины зависимой переменной y . Цель в том, чтобы сделать эту разницу как можно меньше. Однако если эта разница проявляется в наборе данных, то случаи, где величина зависимой переменной выше величины линейной функции, накладываются на те, где величина зависимой переменной ниже величины линейной функции, и они компенсируют друг друга.

Шаг 3

Для определения «отклонения» линейной функции от имеющихся данных возводим в квадрат разницу между линейной функцией $\hat{c}(\mathbf{x})$ и зависимой переменной y ; т.е. находим **квадрат ошибки**. Уменьшение квадрата ошибки путем корректировки веса линейной функции называется обучением по **методу наименьших квадратов**.

Таким образом, добавив к \mathbf{x} в уравнении (1.2) 0-мерность и определив его величину как равную 1, а также добавив w_0 к 0-мерному \mathbf{w} , получим, что функция регрессии будет записываться как внутреннее произведение $(d + 1)$ -мерного вектора (1.3).

$$\hat{c}(\mathbf{x}) = \mathbf{w}^T \mathbf{x}. \quad (1.3)$$

Шаг 4

Оценим коэффициенты \mathbf{w} этого уравнения, используя обучающие данные $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$. Постараемся максимально возможно уменьшить разницу между значениями зависимой переменной y и функции $\hat{c}(\mathbf{x})$, рассчитанной по уравнению 1.3. Ошибка определяется значением коэффициентов \mathbf{w} в уравнении (1.3) обозначив ее $E(\mathbf{w})$, получим следующее уравнение:

$$E(\mathbf{w}) = \sum_{i=1}^n (y_i - \hat{c}(\mathbf{x}_i))^2, \quad (1.4)$$

$$= \sum_{i=1}^n (y_i - \mathbf{w}^T \mathbf{x}_i)^2. \quad (1.5)$$

Шаг 5

Чтобы избавиться от трудоемких вычислений суммы, представим независимые переменные матрицами, а зависимые – векторами. Обозначим матрицу, имеющую n позиций по вертикали, которая получилась путем транспонирования независимой переменной \mathbf{x} d -мерного вектора-столбца как \mathbf{X} , \mathbf{y} – вектор-столбец величин независимой переменной y , а \mathbf{w} – вектор-столбец коэффициентов.

В итоге отклонение примет следующий вид:

$$E(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}). \quad (1.6)$$

Чтобы минимизировать отклонение, нужно найти такие величины коэффициентов \mathbf{w} , при которых производная функции ошибки равняется 0, то есть:

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\mathbf{w}) = 0, \quad (1.7)$$

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (1.8)$$

где \mathbf{A}^{-1} – матрица, обратная матрице \mathbf{A} .



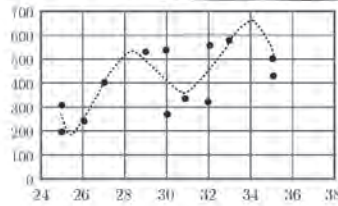
Таким образом, мы можем аналитически найти по обучающим данным веса \mathbf{w} при помощи минимизации суммы квадратов ошибки. При подстановке \mathbf{w} в уравнение (1.3) получится функция линейной регрессии $\hat{c}(\mathbf{x})$.

1.4. РЕГУЛЯРИЗАЦИЯ РЕЗУЛЬТАТА

ЕСЛИ КОЭФФИЦИЕНТЫ ЛИНЕЙНЫЕ, ТО ВЕСА МОГУТ БЫТЬ РАССЧИТАНЫ ТАКИМ ЖЕ ОБРАЗОМ, ДАЖЕ ЕСЛИ ЭТО УРАВНЕНИЯ ВЫСОКОГО ПОРЯДКА. ТАК МОЖНО ПРОВОДИТЬ ОБУЧЕНИЕ ПРИ ПОМОЩИ САМЫХ СЛОЖНЫХ УРАВНЕНИЙ РЕГРЕССИИ.



ЕСЛИ Я ЗНАЮ ЧИСЛЕННЫЙ ВЕС, ТО ОН МОЖЕТ БЫТЬ ПРОСТО ПОДСТАВЛЕН В ЭТУ ФОРМУЛУ?



НЕ СОВСЕМ ТАК. В ЭТОЙ ФОРМУЛЕ ДАЖЕ ПРИ НЕБОЛЬШОМ ИЗМЕНЕНИИ ДАННЫХ НА ВХОДЕ РЕЗУЛЬТАТЫ НА ВЫХОДЕ МОГУТ СИЛЬНО ОТЛИЧАТЬСЯ, И НЕЛЬЗЯ БУДЕТ ПОЛУЧИТЬ ХОРОШИЙ РЕЗУЛЬТАТ ПРИ ИСПОЛЬЗОВАНИИ ДАННЫХ, ОТЛИЧНЫХ ОТ ОБУЧАЮЩИХ.

СЛОЖНОВАТО. А КАК СДЕЛАТЬ ХОРОШИЙ ПРОГНОЗ?



ЕСЛИ НЕБОЛЬШОЕ ИЗМЕНЕНИЕ КАКОГО-ЛИБО ПАРАМЕТРА НА ВХОДЕ ДАЕТ БОЛЬШОЕ ИЗМЕНЕНИЕ НА ВЫХОДЕ, ЭТО ОЗНАЧАЕТ, ЧТО КОЭФФИЦИЕНТ УРАВНЕНИЯ СЛИШКОМ ВЕЛИК, ПОЭТОМУ ЕГО НЕОБХОДИМО УМЕНЬШИТЬ.

ТАК БЫСТРО?

НО, ВПРОЧЕМ, ЕСТЬ И ДРУГАЯ ТОЧКА ЗРЕНИЯ НА ВЕС, В СЛУЧАЯХ, КОГДА ПРАВИЛЬНОСТЬ ПРОГНОЗА ВАЖНЕЕ, ЧЕМ ОБЪЯСНИМОСТЬ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ.



ОБЪЯСНИМОСТЬ ВАЖНЕЕ ТОЧНОСТИ?

С НАВИТЫМ РТОМ



ДОПУСТИМ, ВМЕСТО ТОГО ЧТОБЫ ПРОГНОЗИРОВАТЬ, КАКИЕ ФАКТОРЫ ПОВЛИЯЮТ НА КАЧЕСТВО ТОВАРА, НАДО НАЙТИ ТЕ, КОТОРЫЕ ВЛИЯЮТ НА НЕГО БОЛЬШЕ ВСЕГО.



ТАК МОЖНО НАЙТИ ТЕ ХАРАКТЕРИСТИКИ, КОТОРЫЕ ДЕЛАЮТ ТОВАР БРАКОВАННЫМ.

Пирожок
Составляющие:
- мука
- начинка
- сахар

→ Что влияет больше?

ЕСЛИ ГОВОРИТЬ КОНКРЕТНЕЕ, МОЖНО ПОДАСТАВИТЬ В КАЧЕСТВЕ ВЕСА ПЕРЕМЕННЫХ В ФОРМУЛУ ЛИНЕЙНОЙ РЕГРЕССИИ 0 И ПОСМОТРЕТЬ, ИЗМЕНИТСЯ ЛИ КОЛИЧЕСТВО ИЗМЕРЕНИЙ.



ДРУГИМИ СЛОВАМИ, СЛЕДУЕТ НАЙТИ СПОСОБ, ЧТОБЫ КОЭФФИЦИЕНТ w В УРАВНЕНИИ ЛИНЕЙНОЙ РЕГРЕССИИ УМЕНЬШИЛСЯ, ЕСЛИ ВЕЛИЧИНА ЕГО БОЛЬШАЯ, ИЛИ ЖЕ СТАЛ РАВЕН НУЛЮ.

И ЭТОТ СПОСОБ НАЗЫВАЕТСЯ РЕГУЛЯРИЗАЦИЕЙ.

Регуляризация

ОН НУЖЕН, ЧТОБЫ КОЭФФИЦИЕНТ НЕ БЫЛ СЛИШКОМ БОЛЬШИМ...

МЕТОД РЕГУЛЯРИЗАЦИИ, ПРИ КОТОРОМ МЫ УМЕНЬШАЕМ БОЛЬШИЕ КОЭФФИЦИЕНТЫ, НАЗЫВАЕТСЯ РИДЖ-РЕГРЕССИЕЙ,

А ЧТОБЫ УВЕЛИЧИТЬ КОЛИЧЕСТВО ВЕЛИЧИН, РАВНЫХ НУЛЮ, ИСПОЛЬЗУЕТСЯ ЛАССО-РЕГРЕССИЯ.

РИДЖ-РЕГРЕССИЯ

ЛАССО-РЕГРЕССИЯ

РЕГУЛЯРИЗАЦИЯ ОСУЩЕСТВЛЯЕТСЯ ДОБАВЛЕНИЕМ ДОПОЛНИТЕЛЬНОГО ЧЛЕНА К УРАВНЕНИЮ ОШИБКИ.

НАЧНЕМ С ОБЪЯСНЕНИЯ РИДЖ-РЕГРЕССИИ.

МЫ ДОБАВЛЯЕМ ДОПОЛНИТЕЛЬНЫЙ ЧЛЕН, КВАДРАТ ПАРАМЕТРА w .

Чтобы уменьшить это...
...надо увеличить это

$$E(w) = (y - Xw)^T (y - Xw) + \alpha w^T w$$

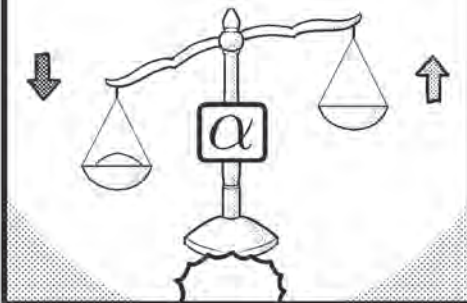
Регулируем баланс

А α ОТКУДА ВЗЯЛАСЬ?

Если величина веса слишком маленькая...

...то она далека от правильной

α - ЭТО ВЕС ДОПОЛНИТЕЛЬНОГО ЧЛЕНА РЕГУЛЯРИЗАЦИИ. ЕСЛИ ПАРАМЕТР БОЛЬШОЙ, ТО ЭФФЕКТ РЕГУЛЯРИЗАЦИИ СТАНОВИТСЯ ВАЖНЕЕ ЭФФЕКТИВНОСТИ, ЕСЛИ МАЛЕНЬКИЙ, ТО ЭФФЕКТИВНОСТЬ СТАНОВИТСЯ ВАЖНЕЕ.



ИСПОЛЬЗУЯ РИДЖ-РЕГРЕССИЮ, МЫ НАХОДИМ ВЕЛИЧИНУ w , КОГДА ГРАДИЕНТ ПО w ФУНКЦИИ ОШИБКИ РАВЕН 0, КАК И В СЛУЧАЕ НАХОЖДЕНИЯ ЭТОГО ПАРАМЕТРА МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ.

$$w = (X^T X + \alpha I)^{-1} X^T y$$



I - единичная матрица.

КСТАТИ, А ПОЧЕМУ РИДЖ-РЕГРЕССИЯ ТАК НАЗЫВАЕТСЯ?



А ТЕПЕРЬ ПОГОВОРИМ О РЕГРЕССИИ "ЛАССО". ЭТО РЕГУЛЯРИЗАЦИЯ, ПРИ КОТОРОЙ w СТАНОВИТСЯ АБСОЛЮТНОЙ ВЕЛИЧИНОЙ.



RIDGE ОЗНАЧАЕТ ГРЕБЕНЬ ГОРЫ, И ЕДИНИЧНАЯ МАТРИЦА НА НЕГО ПОХОЖА*.

* Есть другие версии.

ХМ...

RIDGE - ЭТО КВАДРАТ w , ЛАССО - ЭТО АБСОЛЮТНАЯ ВЕЛИЧИНА w .

А ЧТО ТАКОЕ ЛАССО?

ЛАССО - ЭТО ПЕТЛЯ ДЛЯ ЛОВЛИ КОГО-НИБУДЬ.



КАК Я И ГОВОРИЛА, РИДЖ-РЕГРЕССИЯ - ЭТО РЕГУЛЯРИЗАЦИЯ ДЛЯ УМЕНЬШЕНИЯ ВЕЛИЧИНЫ ПАРАМЕТРА.



ВРОДЕ ЭТО СЛОВО ИСПОЛЬЗУЮТ, КОГДА ГОВОРЯТ О КОВЕОЯХ.

ПРЕДСТАВЬ, ЧТО ВО МНОЖЕСТВО ПАРАМЕТРОВ КИДАЮТ ЛАССО И ВЫБИРАЮТ САМЫЕ МАЛЕНЬКИЕ ИЗ НИХ.

ИЗНАЧАЛЬНО ЭТО АББРЕВИАТУРА ФРАЗЫ **LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR**.

НАДО ЖЕ.

УРАВНЕНИЕ ОЦЕНКИ ОШИБОК РЕГРЕССИИ "ЛАССО" ВЫГЛЯДИТ...

$$E(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + \alpha \sum_{i=1}^n |w_j|$$

...ТАК:

ПОСКОЛЬКУ w_0 - ПОСТОЯННОЕ СЛАГАЕМОЕ В УРАВНЕНИИ, ЕГО ВЕЛИЧИНА НЕ ПОВЛИЯЕТ НА ВЕЛИЧИНУ УРАВНЕНИЯ РЕГРЕССИИ, И ЕГО ОБЫЧНО НЕ РЕГУЛЯРИЗИРУЮТ. ЗДЕСЬ ЕСЛИ ВЕС ДОПОЛНИТЕЛЬНОГО ЧЛЕНА

УВЕЛИЧИВАЕТСЯ, РАСТЕТ ЧИСЛО ВЕЛИЧИН С ВЕСОМ, РАВНЫМ 0.

А КАК МОЖНО ОБЪЯСНИТЬ РЕГРЕССИЮ "ЛАССО"?

Функция абсолютного значения

Квадратичная функция

ПОСКОЛЬКУ ТУДА ВХОДИТ АБСОЛЮТНАЯ ВЕЛИЧИНА, НЕДИФФЕРЕНЦИРУЕМАЯ В ТОЧКЕ НАЧАЛА КООРДИНАТ, НЕЛЬЗЯ НАЙТИ ЗНАЧЕНИЕ АНАЛИТИЧЕСКИ, ИСПОЛЬЗУЯ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ, ПОЭТОМУ ВЕРХНИЙ ПРЕДЕЛ/МАКСИМУМ ДОПОЛНИТЕЛЬНОГО ЧЛЕНА РЕГУЛЯРИЗАЦИИ ОГРАНИЧИВАЕТСЯ ДИФФЕРЕНЦИРУЕМОЙ КВАДРАТИЧНОЙ ФУНКЦИЕЙ. БЫЛ ПРЕДЛОЖЕН МЕТОД, ЧТОБЫ МНОГОКРАТНО ОБНОВЛЯТЬ ЕЕ ПАРАМЕТРЫ С ЦЕЛЬЮ УМЕНЬШЕНИЯ ОШИБКИ.

С ПОМОЩЬЮ МЕТОДА "ЛАССО" МОЖНО ПРОРЕДИТЬ НЕЗАВИСИМЫЕ ПЕРЕМЕННЫЕ С ВЕСАМИ, НЕ РАВНЫМИ НУЛЮ, И НАЙТИ ТЕ, КОТОРЫЕ ОКАЗЫВАЮТ ВЛИЯНИЕ.

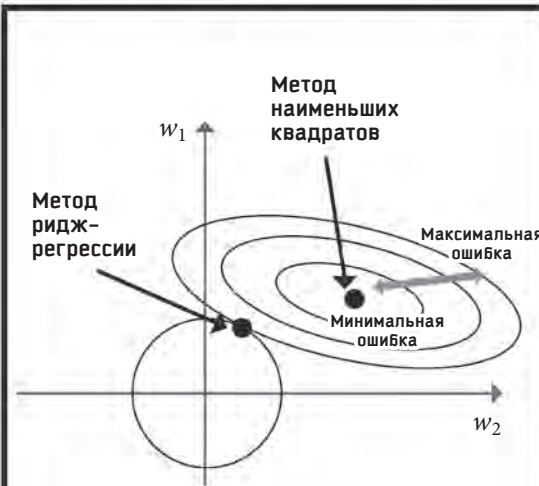


УДОБНО.

А ТЕПЕРЬ Я ОБЪЯСНЮ, ЧЕМ РИДЖ-РЕГРЕССИЯ ОТЛИЧАЕТСЯ ОТ РЕГРЕССИИ "ЛАССО".



ПОЖАЛУЙСТА!



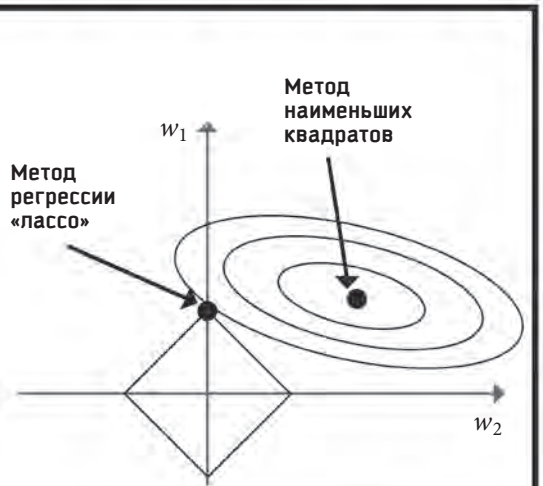
Метод ридж-регрессии

Метод наименьших квадратов

Максимальная ошибка

Минимальная ошибка

w_2



Метод регрессии «лассо»

Метод наименьших квадратов

w_2

КАК ПОКАЗАНО НА РИСУНКЕ, ПРИ ИСПОЛЬЗОВАНИИ РИДЖ-РЕГРЕССИИ ОГРАНИЧЕНИЕ ДИАПАЗОНА ПАРАМЕТРОВ ОКРУЖНОСТЬЮ (ОБЩИЙ СЛУЧАЙ d -МЕРНОЙ ГИПЕРСФЕРЫ) НЕ ПОЗВОЛЯЕТ КАЖДОМУ ВЕСУ ПРИНИМАТЬ БОЛЬШОЕ ЗНАЧЕНИЕ. В ОБЩЕМ СЛУЧАЕ ТОЧКА КАСАНИЯ ИЗОЛИНИИ ФУНКЦИИ ОШИБКИ ЯВЛЯЕТСЯ ТОЧКОЙ НА ОКРУЖНОСТИ, КОТОРАЯ ЯВЛЯЕТСЯ ЗНАЧЕНИЕМ ВЕСА ПАРАМЕТРА.

ПОЭТОМУ ВЕЛИЧИНА ПАРАМЕТРА УМЕНЬШАЕТСЯ.

А В СЛУЧАЕ РЕГРЕССИИ "ЛАССО", ПРИ УСЛОВИИ ЧТО ОПРЕДЕЛЕНА СУММА ПАРАМЕТРОВ, ДИАПАЗОН ПАРАМЕТРОВ ОГРАНИЧЕН ОБЛАСТЬЮ (РОМБОМ), УГЛЫ КОТОРОГО ЛЕЖАТ НА КАЖДОЙ ОСИ, КАК ПОКАЗАНО НА РИСУНКЕ.

И ОДИН ИЗ УГЛОВ РОМБА КАСАЕТСЯ ИЗОЛИНИИ ФУНКЦИИ ОШИБКИ.

КАЖЕТСЯ, В УГЛАХ БОЛЬШИНСТВО ПАРАМЕТРОВ СТАНОВЯТСЯ РАВНЫМИ 0.

ЭТО ВЛИЯНИЕ РЕГУЛЯРИЗАЦИИ РЕГРЕССИИ "ЛАССО".



Для начала загрузим библиотеку. В scikit-learn подготовлено несколько наборов данных, выберем методы из пакета datasets. Для регрессии это линейная регрессия, ридж-регрессия и регрессия «лассо».

```
from sklearn.datasets import load_boston
from sklearn.linear_model import LinearRegression, Ridge, Lasso
```



В качестве данных для анализа мы возьмем 13 параметров из стандартной выборки boston dataset, куда входят уровень преступности, количество комнат, географическое положение и прочие данные, связанные с недвижимостью.



Атрибут data экземпляра boston, который создан при помощи приведенного ниже кода, является матрицей, в которой признаковое описание объекта располагается в виде столбцов (для 13-мерного признака будет 506 векторов-строк), а атрибут target будет введен в качестве вектора-столбца – цены каждого свойства.



Можно показать детали данных boston с помощью атрибута desc функцией print (boston.DESCR).

```
boston = load_boston()
X = boston.data
y = boston.target
```




Код в `scikit-learn` удобен для обучения. А теперь используем учебный набор данных.

```
lr1 = LinearRegression()
```



В этом экземпляре можно вызвать метод `fit`, который выполняет обучение с набором признаков описаний X и точной информацией y в качестве аргументов.

```
lr1.fit(X, y)
```



Когда получено уравнение линейной регрессии, можно узнать прогнозируемое значение, которое в качестве аргумента имеет 13-мерный вектор x , с помощью метода `predict`, который выведет прогнозируемое значение.



А теперь попробуем провести регуляризацию. Для начала используем сумму квадратов и коэффициенты формулы линейной регрессии, которые мы только что разобрали.

```
print("Linear Regression")
for f, w in zip(boston.feature_names, lr1.coef_) :
    print("{0:7s}: {1:6.2f}".format(f, w))
print("coef = {0:4.2f}".format(sum(lr1.coef_**2)))
```

```
Linear Regression
CRIM   : -0.11
ZN     :  0.05
INDUS  :  0.02
CHAS   :  2.69
NOX    : -17.80
RM     :  3.80
AGE    :  0.00
DIS    : -1.48
```

```
RAD : 0.31
TAX : -0.01
PTRATIO: -0.95
B : 0.01
LSTAT : -0.53
coef = 341.86
```



Попробуем также провести ридж-регрессию. Поскольку в данные входят X и y , лучше всего начать с постройки экземпляра, по которому можно провести обучение.

Если есть параметр, который нужно указать, он задается в качестве аргумента экземпляра в формате «имя параметра = значение». Вес α дополнительного параметра регуляризации примем равным 10,0.

```
lr2 = Ridge(alpha=10.0)
lr2.fit(X, y)
print("Ridge")
for f, w in zip(boston.feature_names, lr2.coef_) :
    print("{0:7s}: {1:6.2f}".format(f, w))
print("coef = {0:4.2f}".format(sum(lr2.coef_**2)))
```

```
Ridge
CRIM : -0.10
ZN : 0.05
INDUS : -0.04
CHAS : 1.95
NOX : -2.37
RM : 3.70
AGE : -0.01
DIS : -1.25
RAD : 0.28
TAX : -0.01
PTRATIO: -0.80
B : 0.01
LSTAT : -0.56
coef = 25.73
```



Мы видим, что сумма квадратов коэффициентов абсолютно мала. А теперь применим регрессию «лассо». Вес α дополнительного параметра регуляризации примем равным 2,0 и заметим, что несколько коэффициентов равны 0.

```
lr3 = Lasso(alpha=2.0)
lr3.fi t(X, y)
print("Lasso")
for f, w in zip(boston.feature_names, lr3.coef_) :
    print("{0:7s}: {1:6.2f}". format(f, w))
print("coef = {0:4.2f}".format(sum(lr3.coef_**2)))
```

```
Lasso
CRIM   : -0.02
ZN     :  0.04
INDUS  : -0.00
CHAS   :  0.00
NOX    : -0.00
RM     :  0.00
AGE    :  0.04
DIS    : -0.07
RAD    :  0.17
TAX    : -0.01
PTRATIO: -0.56
B      :  0.01
LSTAT  : -0.82
coef = 1.02
```