

Посвящается моему сыну Арану

Краткое оглавление

Часть I. Причинно-следственный анализ	
и роль искажающих факторов	27
1 ■ <i>Введение в причинность.....</i>	<i>29</i>
2 ■ <i>Первые шаги: работа с искажающими факторами.....</i>	<i>60</i>
3 ■ <i>Применение причинно-следственного анализа</i>	<i>102</i>
4 ■ <i>Как машинное обучение и причинно-следственный анализ могут помочь друг другу</i>	<i>123</i>
Часть II. Формула корректировки на практике	161
5 ■ <i>Поиск сопоставимых случаев с использованием меры склонности.....</i>	<i>163</i>
6 ■ <i>Расчет прямых и косвенных эффектов с помощью линейных моделей.....</i>	<i>198</i>
7 ■ <i>Работа со сложными графами</i>	<i>228</i>
8 ■ <i>Расширенные инструменты в библиотеке DoubleML....</i>	<i>276</i>
Часть III. Другие стратегии помимо формулы корректировки	305
9 ■ <i>Инструментальные переменные.....</i>	<i>307</i>
10 ■ <i>Схема оценки потенциальных исходов</i>	<i>332</i>
11 ■ <i>Эффекты событий во времени.....</i>	<i>350</i>

Оглавление

Предисловие от издательства	14
Предисловие	15
Благодарности	16
О книге.....	18
Об авторе.....	24
Об иллюстрации на обложке	25

ЧАСТЬ I. Причинно-следственный анализ и роль искажающих факторов 27

1 Введение в причинность 29

1.1. Как работает причинно-следственный анализ.....	31
1.1.1. Шаг 1: определение типа данных	31
1.1.2. Шаг 2: изучение задачи	32
1.1.3. Шаг 3: создание модели	32
1.1.4. Шаг 4: представление модели другим.....	33
1.1.5. Шаг 5: применение методов причинно-следственного анализа.....	33
1.2. Различия между причинно-следственными моделями и прогностическими моделями машинного обучения.....	35
1.3. Экспериментальные исследования.....	37
1.3.1. Мотивирующий пример: развертывание нового веб-сайта	37
1.3.2. А/В-тестирование.....	40
1.3.3. Рандомизированные контролируемые исследования	41
1.3.4. Шаги проведения А/В-тестирования.....	42
1.3.5. Ограничения А/В-тестирования и RCT	44
1.4. Наблюдательные исследования.....	45
1.4.1. Моделирование синтетических данных	47
1.4.2. Причинно-следственные связи при наличии искажающих факторов.....	49
1.5. Обзор основных статистических концепций	51
1.5.1. Эмпирические распределения и распределения сгенерированных данных	51
1.5.2. Условные вероятности и ожидания	53
1.6. Для дополнительного чтения	57
1.7. Контрольные вопросы.....	58
Итоги.....	59

2 Первые шаги: работа с искажающими факторами 60

2.1. Основы причинно-следственного анализа и парадокс Симпсона	62
2.1.1. Чем обусловлена эта проблема?.....	64
2.1.2. Развивайте интуицию: как исправить проблему.....	66
2.1.3. Решение парадокса Симпсона	67
2.2. Обобщение на другие задачи	69

2.2.1. Описание проблемы с помощью графа	70
2.2.2. Формулирование того, что хотелось бы узнать	70
2.2.3. Поиск способа вычисления причинно-следственной связи	71
2.2.4. Формулирование того, что хотелось бы узнать: язык вмешательств	71
2.2.5. Поиск способа вычисления причинно-следственной связи: формула корректировки	74
2.2.6. Какие результаты дает метод лечения в каждой ситуации? Предположение о положительности	76
2.3. Вмешательства и RCT	78
2.4. Первая встреча со структурным подходом	79
2.4.1. Моделирование примера почечнокаменной болезни	81
2.4.2. Вмешательства в структурном подходе	83
2.5. Когда применять формулу корректировки	85
2.5.1. RCT или A/B-тестирование	86
2.5.2. Искажающие факторы	87
2.5.3. Ненаблюдаемые искажающие факторы	88
2.5.4. Медиаторы	89
2.5.5. Множество искажающих факторов	90
2.5.6. Переменные, прогнозирующие результат	91
2.5.7. Переменные, прогнозирующие выбор метода лечения	93
2.5.8. Условное вмешательство	93
2.5.9. Объединение всех предыдущих ситуаций	95
2.5.10. Обобщение различий между вмешательством и применением формулы корректировки	96
2.6. Итак, каков план?	97
2.7. Основные уроки главы	99
2.8. Контрольные вопросы	100
Итоги	101

3 *Применение причинно-следственного анализа* 102

3.1. Когда и зачем использовать графы в причинно-следственном анализе	103
3.2. Шаги по формулированию задачи с использованием графов	105
3.2.1. Составьте список всех переменных	106
3.2.2. Создайте свой граф	108
3.2.3. Сформулируйте свои предположения	112
3.2.4. Сформулируйте свои цели	113
3.2.5. Проверьте предположение о положительности	114
3.3. Другие примеры	115
3.3.1. Рекомендательные системы	115
3.3.2. Ценообразование	118
3.3.3. Моделирование	118
3.4. Для дальнейшего чтения	121
3.5. Контрольные вопросы	121
Итоги	121

4 *Как машинное обучение и причинно-следственный анализ могут помочь друг другу* 123

4.1. Что дает обучение с учителем?	126
4.1.1. Когда следует использовать причинно-следственный анализ, а когда обучение с учителем?	128

4.1.2. Цель аппроксимации данных.....	129
4.1.3. Когда будущее и прошлое имеют одинаковое поведение.....	131
4.1.4. Когда причинно-следственный анализ и обучение с учителем совпадают?.....	133
4.1.5. Ошибка прогнозирования – ложный друг	134
4.1.6. Проверка вмешательств.....	140
4.2. Как обучение с учителем участвует в причинно-следственном анализе?	141
4.2.1. Эмпирические распределения и распределения сгенерированных данных в формуле корректировки	143
4.2.2. Гибкость формулы корректировки	144
4.2.3. Формула корректировки для непрерывных распределений	145
4.2.4. Алгоритмы расчета формулы корректировки	145
4.2.5. Перекрестное обучение: предотвращение переобучения	149
4.3. Другие применения причинно-следственного анализа в машинном обучении	153
4.3.1. Обучение с подкреплением.....	153
4.3.2. Справедливость	155
4.3.3. Ложные корреляции.....	155
4.3.4. Обработка естественного языка.....	156
4.3.5. Объяснимость.....	156
4.4. Для дальнейшего чтения.....	156
4.5. Контрольные вопросы.....	158
Итоги.....	158

ЧАСТЬ II. Формула корректировки на практике161

5 Поиск сопоставимых случаев с использованием меры склонности.....163

5.1. Знакомство с мерой склонности	166
5.1.1. Поиск соответствий для оценки причинно-следственных связей.....	167
5.1.2. Но есть ли соответствие?.....	168
5.1.3. Почему сопоставление может быть трудным	169
5.1.4. Как меры склонности можно использовать для расчета АТЕ....	171
5.2. Основные понятия мер склонности	172
5.2.1. С какими случаями мы работаем?	173
5.2.2. Что такое мера склонности?	176
5.2.3. Предположение о положительности – это... предположение	176
5.3. Оценки меры склонности на практике	177
5.3.1. Подготовка данных.....	177
5.3.2. Вычисление меры склонности.....	178
5.3.3. Оценка предположения о положительности	181
5.3.4. Вычисление АТЕ на основе мер склонности.....	187
5.4. Вычисление корректировки меры склонности: упражнение.....	194
5.4.1. Шаги упражнения	195
5.5. Для дальнейшего чтения.....	196
5.6. Контрольные вопросы.....	196
Итоги.....	197

6 Расчет прямых и косвенных эффектов с помощью линейных моделей198

6.1. Оценка причинно-следственных связей	
с помощью линейных моделей	201
6.1.1. Моделирование задачи ценообразования: знакомство	201
6.1.2. Прямые и косвенные эффекты.....	207
6.2. Изучение причинно-следственной динамики с помощью	
линейных моделей	214
6.2.1. Аналогия с газом, текущим по трубам.....	215
6.2.2. Как корреляция распространяется в графе	215
6.2.3. Расчет причинно-следственной связи и корреляции	
по коэффициентам при стрелках.....	220
6.2.4. Линейные модели и оператор «do»	222
6.3. Контрольные вопросы	226
Итоги.....	226

7

Работа со сложными графами.....228

7.1. Изменение корреляции между двумя переменными	
с учетом третьей.....	232
7.1.1. Пример условной независимости времени прихода	
на работу.....	233
7.1.2. Математический пример условной независимости	234
7.1.3. Разбиение причинно-следственной модели	
на независимые модули.....	235
7.1.4. Кирпичики DAG: факторизация	
распределений вероятностей.....	240
7.1.5. Что такое d-разделение?	247
7.1.6. Определение d-разделения.....	251
7.2. Критерий обходного пути	253
7.2.1. Важность критерия обходного пути	258
7.3. Хорошие и плохие наборы корректируемых переменных.....	261
7.3.1. Хорошие наборы корректируемых переменных.....	261
7.3.2. Нейтральные наборы корректируемых переменных	262
7.3.3. Плохие наборы корректируемых переменных.....	262
7.4. Возвращаясь к предыдущим главам.....	264
7.4.1. Эффективные наборы корректируемых переменных.....	264
7.4.2. Оценка склонности.....	268
7.4.3. И снова: не включайте переменные в модель	
только потому, что они делают ее более точной	269
7.4.4. Следует ли делать поправку на доход?	269
7.5. Дополнительный инструмент для выявления	
причинно-следственных связей: do-исчисление	272
7.6. Для дальнейшего чтения.....	273
7.7. Контрольные вопросы.....	274
Итоги.....	275

8

Расширенные инструменты в библиотеке DoubleML.....276

8.1. Двойное машинное обучение.....	279
8.1.1. Теорема FWL: предшественница DML	281
8.1.2. Нелинейные модели в DML	285
8.1.3. DML на практике.....	290
8.1.4. Гетерогенные эффекты воздействия	294
8.2. Доверительные интервалы	296

8.2.1. Моделирование новых наборов данных с помощью бутстрэппинга	297
8.2.2. Аналитические формулы вычисления доверительных интервалов.....	298
8.3. Оценки с двойной надежностью	300
8.3.1. AIPW на практике.....	302
8.4. Для дальнейшего чтения.....	302
8.5. Контрольные вопросы.....	303
Итоги.....	303

ЧАСТЬ III. Другие стратегии помимо

формулы корректировки305

9 *Инструментальные переменные* 307

9.1. Знакомство с IV на примере.....	309
9.1.1. Граф примера.....	311
9.1.2. Предположения IV.....	312
9.1.3. Инструментальные переменные и RCT	314
9.2. Оценка причинно-следственной связи с помощью IV.....	315
9.2.1. Применение IV с линейными моделями	315
9.2.2. Применение IV с частично линейными моделями.....	318
9.2.3. Альтернативная формула для метода IV.....	319
9.2.4. Отсутствие общей формулы для обобщенного графа IV.....	320
9.3. Инструментальные переменные на практике.....	320
9.3.1. Двухэтапный алгоритм наименьших квадратов (2SLS)	322
9.3.2. Слабые инструменты.....	324
9.3.3. IV и DoubleML	327
9.4. Ссылки	330
9.5. Контрольные вопросы.....	331
Итоги.....	331

10 *Схема оценки потенциальных исходов*332

10.1. Что такое потенциальный исход?	333
10.1.1. Индивидуальные исходы	334
10.1.2. Коллективные исходы.....	336
10.1.3. Причинно-следственные эффекты	337
10.1.4. Предположения о потенциальных исходах	338
10.2. Связь метода потенциальных исходов с методом графов.....	339
10.2.1. Первый закон причинно-следственного анализа	339
10.2.2. Выражение предположений PO с помощью DAG	340
10.2.3. Контрфактуалы	341
10.3. Формула корректировки с потенциальными исходами	343
10.4. Инструментальные переменные с потенциальными исходами.....	347
10.5. Контрольные вопросы.....	348
Итоги.....	349

11 *Эффекты событий во времени*350

11.1. Какие типы данных будут использоваться?	354
11.2. Метод разрывной регрессии.....	355
11.2.1. Моделирование данных	356
11.2.2. Терминология RDD.....	358
11.2.3. Предположения.....	360

11.2.4. Оценка эффекта	360
11.2.5. RDD на практике	361
11.3. Синтетический контроль	373
11.3.1. Моделирование данных	376
11.3.2. Терминология метода синтетического контроля	377
11.3.3. Предположения	378
11.3.4. Оценка эффекта	379
11.3.5. Синтетический контроль на практике	380
11.3.6. Выбор периодов обучения и прогнозирования	380
11.4. Метод сравнения разностей	386
11.4.1. Моделирование данных	388
11.4.2. Терминология DiD	389
11.4.3. Предположения	392
11.4.4. Оценка эффекта	394
11.4.5. Практика	395
11.5. Контрольные вопросы	401
11.6. Сравнение методов	401
11.7. Ссылки	402
Итоги	402
Приложение А. Математика, лежащая в основе формулы коррективки	403
Приложение В. Решения упражнений в главе 2	407
В.1. Решение парадокса Симпсона для метода лечения В	407
В.2. Наблюдать и делать – это не одно и то же	407
В.2.1. Решение	408
В.3. Что нужно скорректировать?	409
В.3.1. RCT	409
В.3.2. Искажающие факторы	410
В.3.3. Ненаблюдаемые искажающие факторы	412
В.3.4. Медиаторы	413
В.3.5. Переменные, прогнозирующие результат	414
Приложение С. Техническая лемма о мерах склонности	416
Приложение D. Доказательство двойной надежности оценки \widehat{ATE}_{airpw}	420
D.1. Свойство двойной надежности по отношению к T-learner	420
D.2. Свойство двойной надежности относительно обратного взвешивания вероятности	421
Приложение Е. Техническая лемма для альтернативной формулы инструментальной переменной	423
Приложение F. Доказательство формулы неидеального соответствия инструментальной переменной	424
Предметный указатель	427

Предисловие от издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Предисловие

Впервые проблема причинно-следственного анализа привлекла мое внимание в 2016 году, когда я прочитал статью о причинности и совершенно ничего не понял. Заинтересовавшись этой экзотической темой, я прочитал работы Джуды Перла (Judea Pearl). Поначалу я не думал, что с помощью математических и статистических инструментов можно многого добиться в моделировании причинности. Но, продолжая читать, я понял, что причинность имеет большое значение с прикладной точки зрения.

В 2016 году я уже несколько лет работал специалистом по данным. Мне нравилось исследовать методы машинного обучения; это был новый мир, в котором мое техническое образование давало хорошее преимущество. Машинное обучение открыло двери в различные отрасли и компании, позволяя мне наслаждаться своей работой.

Между 2016 и 2018 годами я начал понимать, что прогресс в машинном и особенно в глубоком обучении в значительной мере был достигнут путем проб и ошибок, без глубокого понимания его внутренних механизмов. Основное внимание уделялось вычислительной мощности и программированию, а не моделированию мира. Я не против такого подхода, но он не удовлетворял меня. В то же время я все глубже исследовал область причинно-следственного анализа. С каждым днем я обнаруживал, что все больше разделяю его цели – получение ответа на вопрос «почему» – и его методы, уделяющие большое внимание статистике и математике, но при этом включающие и программирование.

В конце 2018 года я решил взять годовой отпуск, чтобы вплотную заняться темой причинно-следственного анализа. Я понял, что компании часто испытывают трудности с точной оценкой эффекта своих решений, и казалось неизбежным, что они быстро осознают необходимость выявления причинно-следственных связей и создадут вакансии для специалистов в этом направлении. Когда я обсуждал рабочие проблемы со своими коллегами, то заметил, что способен быстрее и точнее понять и проанализировать их проблемы, чем раньше. А на личном уровне изучение причинно-следственного анализа кардинально изменило мой взгляд на мир: я начал находить причинно-следственные связи во многих сферах жизни общества, включая здравоохранение, экономику, журналистику, политику и т. д.

Приятно тратить время на получение новых знаний, которые могут принести мне доход, а также помочь лучше понять мир, в котором мы живем. С 2018 года я объяснял основы причинно-следственного анализа на многих семинарах, занятиях, в статьях, а теперь и в этой книге! Каждый раз, когда я рассказываю об этой теме, я испытываю то же волнение, что и тогда, когда впервые ее открыл.

Благодарности

Начав писать эту книгу, я думал, что мне понадобится год. Какой же я был наивный! Работа над книгой заняла три года, и я очень горжусь и доволен результатом. Я думаю, что мне удалось найти точный баланс между теорией и практикой, между интуицией и формализмом, что не всегда легко сделать.

Я благодарю свою жену Терезу за поддержку всех моих сумасшедших идей и проектов. Я люблю тебя (и далеко не только потому, что ты поддерживала меня во время работы над этой книгой!). Я также благодарю мою маму Ану, отца Кике и сестру Джулию за постоянную поддержку на протяжении всей моей жизни. Моя потребность понять, почему все так, как есть, и что мы можем сделать, чтобы изменить это, во многом воспитана ими.

Огромное спасибо издательству Manning за то, что вложили время, силы, терпение и деньги в создание этой книги. Из членов команды Manning я хотел бы особо поблагодарить Фрэнсис Лефковиц (Frances Lefkowitz). Без нее эта книга не получилась бы такой, какая она есть. Но не только за это: она также сделала все, чтобы этот долгий писательский труд принес свои плоды. Ее наставничество и точка зрения значительно повлияли на мою манеру общения с другими в повседневной работе. (Сейчас она, вероятно, думает, как бы лучше перефразировать эти предложения!) Работать с вами, Фрэнсис, было истинным удовольствием. Я также хотел бы поблагодарить производственную команду Manning, которая постаралась привести эту книгу к ее окончательному виду.

Не могу не вспомнить также моих друзей и коллег, внесших множество исправлений и предложений для улучшения книги: это Люренс Бадьелла (Llorenç Badiella), Карлос Очоа (Carlos Ochoa), Бартек Скорульский (Bartek Skorulski), Мария Хосе Пелаес (María José Peláez) и Хесус Серкидес (Jesús Cerquides). Большое вам спасибо! Я также очень благодарен рецензентам, чьи предложения помогли сделать эту книгу лучше: это Ади Шавит (Adi Shavit), Алексей Агарков (Alexey Agarkov), Аюш Бихани (Ayush Bihani), Бенедикт Штеммлер (Benedikt Stemmler), Брайан Коколиккио (Brian Cocolicchio), Карлос Айя-Морено (Carlos Aya-Moreno), Эдгар Хасслер (Edgar Hassler), Эзра Шредер (Ezra Schroeder), Грегорио Пикколи (Gregorio Piccoli), Джиллиан Моррисон (Jillian Morrison), Кали Канеко (Kali Kaneko), Каран Гупта (Karan Gupta), Кей Энгельхардт (Kay Engelhardt), Ким Фальк Йоргенсен (Kim Falk Jørgensen), Лара Томпсон (Lara Thompson), Максим Волгин (Maxim Volgin), Микаэль Дотри (Mikael Dautrey), Питер Хенсток (Peter Henstock), Питер Рабинович (Peter Rabinovitch), Салил Атали

(Salil Athalye), Серджио Говони (Sergio Govoni), Шантану Нима (Shantanu Neema), Симона Сгуазца (Simone Sguazza), Судипта Мукерджи (Sudipta Mukherjee), Томас Джозеф Хейман (Thomas Joseph Heiman), Тимоти Лин (Timothy Lin), Тони Дубицкий (Tony Dubitsky) и Тимотеуш Володзко (Tymoteusz Wołodźko). Спасибо также Шону Тейлору (Sean Taylor), Матеусу Факуре (Matheus Facure), Филиппу Баху (Philipp Bach) и Полу Хунермунду (Paul Hünermund) за их поддержку.

Особое спасибо Рамону Наварро (Ramon Navarro) и Эудальду Кампруби (Eudald Camprubi) – основателям стартапа Nuclia в городе Манресе (Каталония) – за то, что поняли мою страсть к причинно-следственным связям и оставили меня в своей команде. Рамон и Эдуард, в глубине души вы знаете, что вашему стартапу было бы лучше в Грасии, Барселона, но все мы знаем, что в этой жизни нет ничего идеального!

Наконец, я выражаю признательность многим исследователям, посвятившим десятилетия продвижению темы причинно-следственных связей и выводу ее на передний план для всех нас, чтобы мы могли продолжить развивать ее. Я хочу поблагодарить Джуду Перла (Judea Pearl), выразив свою благодарность способом, который он бы точно оценил: от противного. Если бы не он, я бы никогда не рискнул войти в мир причинно-следственной связи. Список исследователей, внесших вклад в причинно-следственную связь, велик. Среди них Э. Барейнбойм (E. Bareinboim), расширивший область применения графовых моделей в причинно-следственной связи. Методы причинно-следственного анализа с более статистическим и эконометрическим уклоном были разработаны благодаря таким людям, как Дж. Робинс (J. Robins), Д. Рубин (D. Rubin), В. Черножуков (V. Chernozhukov), Г. Имбенс (G. Imbens), Дж. Ангрист (J. Angrist), М. Дж. ван дер Лаан (M.J. van der Laan), А. Ротницки (A. Rotnitzky) и Э. Х. Кеннеди (E.H. Kennedy). Затем следует Б. Шолкопф (B. Schölkopf) с его коллегами, углубившиеся в исследование связей между причинно-следственной связью и машинным обучением. И это только малая часть, на самом деле таких исследователей очень и очень много.

Эта книга была написана и для начинающих, и для опытных специалистов по данным, для практиков и исследователей машинного обучения, для аналитиков данных, экономистов и статистиков, желающих усовершенствовать процесс принятия решений с использованием результатов наблюдений. Ее цель – дать вам прочную основу для применения методов причинно-следственного анализа в повседневных задачах. Она даст вам интуитивное понимание в выборе правильных инструментов, которое в сочетании с более формальным подходом гарантирует вам уверенность в своих действиях.

Предварительные требования

Для работы с книгой вам понадобятся базовые знания по следующим темам.

- Теория вероятности:
 - основные формулы вероятности, такие как закон полной вероятности и условные вероятности;
 - базовые распределения вероятностей, такие как гауссово и биномиальное;
 - как генерировать случайные числа с помощью компьютера.
- Статистика:
 - линейная и логистическая регрессия;
 - доверительные интервалы;
 - рекомендуется: понимание основ А/В-тестирования и рандомизированных контролируемых испытаний (как выполняется распределение по группам и проверка гипотез).
- Программирование:
 - базовые навыки программирования (чтение/написание простых программ) как минимум на одном языке программирования, например Python, R или Julia.
- Машинное обучение:
 - перекрестная проверка и настройка гиперпараметров;
- рекомендуется: опыт работы с моделями машинного обучения, такими как kNN, случайные леса, бустинг и глубокое обучение.

Структура книги

Книга разделена на три части. Часть I посвящена основам, необходимым для понимания причин и следствий. Здесь вы узнаете, когда применять причинно-следственный анализ; как определенные переменные, известные как искажающие факторы, могут затруднить анализ; и как оценить причинно-следственные связи, устранив их влияние с помощью метода, называемого формулой корректировки.

- Глава 1 представляет два способа причинно-следственного анализа на основе данных: с помощью экспериментов (этот способ еще называют А/В-тестированием, или рандомизированным контролируемым испытанием) или без них. Здесь также объясняются риски анализа неэкспериментальных данных при наличии искажающих факторов.
- Глава 2 представляет формулу корректировки, которая оценивает причинно-следственное влияние в неэкспериментальных данных путем устранения влияния искажающих факторов.
- Глава 3 приводит примеры моделирования анализа с использованием графов.
- Глава 4 объясняет использование машинного обучения для расчета формулы корректировки и показывает, как причинно-следственный анализ может улучшить некоторые аспекты машинного обучения.

В части II рассматриваются реальные проблемы, с которыми можно столкнуться при использовании формулы корректировки из части I.

- Глава 5 рассказывает, как выявить нехватку данных в оцениваемом вами решении.
- Глава 6 объясняет, как оценить причинно-следственные связи непрерывных переменных с помощью линейных моделей.
- Глава 7 рассказывает, как с помощью критерия обходного пути выбрать переменные для включения в анализ, если причинно-следственный граф слишком сложный.
- Глава 8 описывает прием двойного машинного обучения (продвинутый метод оценки причинно-следственных связей) и доверительные интервалы. В главе также объясняется, как их вычислить с помощью пакета DoubleML.

В части III показаны дополнительные методы изучения причин и следствий помимо формулы корректировки.

- Глава 9 представляет метод инструментальных переменных, использующий независимый источник вариации для оценки причинно-следственной связи, и не требует знания каких-либо искажающих факторов, влияющих на результаты.
- Глава 10 обсуждает схему оценки потенциальных исходов – альтернативу графовым причинно-следственным моделям.
- Глава 11 объясняет методы, связанные со временем, часто используемые в экономике, такие как синтетический контроль, метод разрывной регрессии и метод сравнения разностей.

Путь обучения и философия этой книги

Первым делом вы узнаете, как определить, что имеет место причинно-следственная проблема. Не каждый вопрос имеет причинно-следственную связь, иногда просто нужно описать происходящее или предсказать, что произойдет в будущем.

Затем вы пройдете через этапы осмысления причинно-следственных вопросов, включая следующие важные идеи:

- когда нужно провести эксперимент, когда использовать причинно-следственные связи, а когда – машинное обучение;
- использование причинно-следственных диаграмм для представления происходящего в реальном мире;
- использование этих диаграмм для четкого обозначения своих целей, предположений, рисков, на которые вы идете, и что ваши данные могут и не могут вам рассказать;
- проверка наличия всей необходимой для анализа информации (переменных) и выяснение, чего не хватает;
- оценка причинно-следственных связей с использованием статистических методов и методов машинного обучения.

Различные стили обучения

Не все обучаются одинаково: кто-то предпочитает учиться на *примерах*, другие заглядывают в *код*, а кто-то находит более ясными *математические выкладки*. Я включил в книгу все три аспекта. У вас может быть свой предпочтительный метод обучения, но я предлагаю выйти из зоны комфорта и попробовать другие. Если вы легко читаете и понимаете математические формулы, то попробуйте применить методы с использованием синтетических наборов данных. И наоборот, если вам больше нравится исследовать программный код, то попробуйте внимательно прочитать и осмыслить математические доказательства. Чем больше точек зрения на этот материал у вас будет, тем лучше вы его поймете.

Когда я изучаю что-то новое, то следую подходу «сначала думай, потом читай». Далее я перечислю основные его положения для тех, кто захочет попробовать учиться так же, как я.

- Увидев технический термин, я стараюсь самостоятельно вспомнить его определение и только потом заглядываю в книгу.
- Если в книге утверждается какой-то факт, например о том, что искажающие факторы вызывают ложные корреляции, то я обдумываю его смысл, прежде чем прочитать объяснение.
- Когда в книге говорится о математической идее, например о линейной регрессии, я придумываю пример в голове. Благодаря этому мои мысли начинают совпадать с тем, что я читаю.

Поначалу этот подход может показаться немного сложным, потому чтение движется медленнее, но он очень помогает в долгосрочной

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru