

Содержание

От издательства	13
Об авторах	14
Предисловие.....	15
Общепотребительные методы оптимизации, уравнения, символы и акронимы	20
Часть I. Понижение размерности и преобразования	31
Глава 1. Сингулярное разложение (SVD)	32
1.1. Общие сведения.....	33
Определение SVD	34
Вычисление SVD	35
Историческая справка	36
Использование в этой книге и предположения о подготовке читателей.....	37
1.2. Аппроксимация матриц	37
Усечение.....	38
Пример: сжатие изображения.....	38
1.3. Математические свойства и манипуляции	41
Интерпретация с привлечением доминирующих корреляций	41
Метод моментальных снимков.....	43
Геометрическая интерпретация	43
Инвариантность SVD относительно унитарных преобразований	45
Левые унитарные преобразования	46
Правые унитарные преобразования	46
1.4. Псевдообращение, метод наименьших квадратов и регрессия.....	47
Одномерная линейная регрессия	49
Полилинейная регрессия	51
Предостережение	53
1.5. Метод главных компонент (PCA)	53
Вычисление	54
Пример: данные с гауссовым шумом.....	55
Пример: данные о раке яичников	57
1.6. Пример: «собственные лица»	58
1.7. Отсечение и выравнивание.....	64
Оптимальный жесткий порог отсечения.....	64
Важность выравнивания данных.....	68

1.8. Рандомизированное сингулярное разложение	71
Рандомизированная линейная алгебра	71
Рандомизированный алгоритм SVD	72
Пример рандомизированного SVD	75
1.9. Тензорные разложения и N -мерные массивы данных	76
Рекомендуемая литература	81

Глава 2. Преобразование Фурье

и вейвлет-преобразование

2.1. Ряд Фурье и преобразование Фурье	83
Скалярные произведения функций и векторов	83
Ряд Фурье	84
Преобразование Фурье	89
2.2. Дискретное преобразование Фурье (ДПФ) и быстрое преобразование Фурье (БПФ)	92
Дискретное преобразование Фурье	93
Быстрое преобразование Фурье	95
Пример БПФ: фильтрация шума	96
Пример БПФ: спектральные производные	98
2.3. Преобразование дифференциальных уравнений в частных производных	100
Уравнение теплопроводности	101
Одностороннее волновое уравнение	103
Уравнение Бюргера	105
2.4. Преобразование Габора и спектрограмма	107
Дискретное преобразование Габора	108
Пример: сигнал с квадратичной частотной модуляцией	108
Пример: «Патетическая соната» Бетховена	110
Принцип неопределенности	112
2.5. Вейвлеты и многомасштабный анализ	113
Дискретное вейвлет-преобразование	115
2.6. Двумерные преобразования и обработка сигналов	116
Двумерное преобразование Фурье для изображений	116
Двумерное вейвлет-преобразование изображений	119
Рекомендуемая литература	122

Глава 3. Разреженность и сжатие измерений

3.1. Разреженность и сжатие	124
Пример: сжатие изображения	125
Почему сигналы допускают сжатие: просторность пространства изображений	127
3.2. Сжатое измерение	128
Заявление об отказе от ответственности	132
Другие формулировки	133
3.3. Примеры сжатых измерений	133
Норма ℓ_1 и разреженные решения недоопределенной системы	134

Восстановление звукового сигнала по разреженным измерениям	135
3.4. Геометрия сжатия	137
Свойство ограниченной изометрии (RIP)	139
Некогерентность и матрицы измерений.....	139
Плохие измерения	140
3.5. Разреженная регрессия.....	140
Отбрасывание выбросов и робастность	141
Отбор признаков и LASSO-регрессия	142
3.6. Разреженное представление.....	146
3.7. Робастный метод главных компонент (RPCA).....	151
3.8. Разреженное размещение датчиков.....	153
Разреженное размещение датчиков для реконструкции	154
Разреженная классификация	158
Рекомендуемая литература	159

Часть II. МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ

ДАННЫХ

Глава 4. Регрессия и выбор модели.....

4.1. Классическая аппроксимация кривой	163
Методы наименьших квадратов	163
Линия наименьших квадратов.....	166
Линеаризация данных.....	167
4.2. Нелинейная регрессия и градиентный спуск.....	169
Градиентный спуск	170
Метод переменных направлений	175
4.3. Регрессия и уравнение $Ax = b$: переопределенные и недоопределенные системы.....	176
Переопределенные системы	176
Недоопределенные системы	180
4.4. Оптимизация как краеугольный камень регрессии	183
4.5. Парето-фронт и Lex Parsimoniae	188
Переобучение.....	190
4.6. Выбор модели: перекрестная проверка	191
<i>k</i> -групповая перекрестная проверка.....	195
Перекрестная проверка с контролем по <i>p</i> точкам	197
4.7. Выбор модели: информационный критерий	197
Информационные критерии: AIC и BIC	200
Вычисление AIC и BIC.....	201
Рекомендуемая литература	202

Глава 5. Кластеризация и классификация

5.1. Выделение признаков и добыча данных	204
5.2. Обучение с учителем и без учителя.....	210
5.3. Обучение без учителя: кластеризация методом <i>k</i> средних	214
5.4. Иерархическая кластеризация без учителя: дендрограмма.....	219

5.5. Смесовые модели и EM-алгоритм.....	223
5.6. Обучение с учителем и линейные дискриминанты.....	227
5.7. Метод опорных векторов (SVM)	233
Линейный SVM	233
Нелинейный SVM.....	235
Ядерные методы в сочетании с SVM.....	236
5.8. Решающие деревья и случайные леса	238
Случайные леса.....	243
5.9. 10 лучших алгоритмов по версии Data Mining 2008.....	244
Алгоритм k средних	244
EM-алгоритм (смесовые модели)	245
Метод опорных векторов (SVM).....	245
CART (Classification and Regression Tree – дерево классификации и регрессии)	245
Метод k ближайших соседей (kNN).....	246
Наивная байесовская классификация.....	246
AdaBoost (ансамблевое обучение с усилением)	246
C4.5 (ансамблевое обучение решающих деревьев).....	247
Алгоритм Apriori	247
PageRank	247
Рекомендуемая литература	248

Глава 6. Нейронные сети и глубокое обучение

6.1. Нейронные сети: однослойные сети	250
Однослойная сеть.....	252
6.2. Многослойные сети и функции активации	255
6.3. Алгоритм обратного распространения	260
6.4. Алгоритм стохастического градиентного спуска	264
6.5. Глубокие сверточные нейронные сети.....	267
Сверточные слои	268
Пулинговые слои	269
Полносвязные слои	269
Прореживание	270
6.6. Нейронные сети для динамических систем	272
6.7. Разнообразие нейронных сетей	277
Перцептрон	277
Сети прямого распространения (FF)	277
Рекуррентная нейронная сеть (RNN)	279
Автокодировщик (AE).....	279
Марковская цепь (MC)	280
Сеть Хопфилда (HN).....	280
Машина Больцмана (BM)	280
Ограниченная машина Больцмана (RBM)	281
Сеть глубокого доверия (DBN).....	281
Глубокая сверточная нейронная сеть (DCNN).....	281
Антисверточная сеть (DN).....	281
Глубокая сверточная сеть обратной графики (DCIGN).....	282

Порождающая состязательная сеть (GAN)	282
Машина неустойчивых состояний (LSM).....	282
Машина экстремального обучения (ELM)	283
Сеть с эхо-состояниями (ESN)	283
Глубокая остаточная сеть (DRN).....	283
Сеть Кохонена (KN)	284
Нейронная машина Тьюринга (NTM).....	284
Рекомендуемая литература	284

Часть III. ДИНАМИЧЕСКИЕ СИСТЕМЫ И УПРАВЛЕНИЕ.....

285

Глава 7. Динамические системы, управляемые данными.....

286

7.1. Обзор, мотивация и проблемы	287
Динамические системы	287
Цели и проблемы современной теории динамических систем	291
7.2. Разложение по динамическим модам (DMD)	294
Алгоритм DMD	295
Пример и код	300
Расширения, приложения и ограничения.....	300
7.3. Разреженная идентификация нелинейной динамики (SINDy)	308
Нахождение дифференциальных уравнений в частных производных.....	314
Обобщение SINDy на рациональные нелинейности	316
Применение информационного критерия для выбора модели	319
7.4. Теория оператора Купмана	320
Математическая теория оператора Купмана.....	320
Разложение по модам Купмана и конечные представления.....	324
Примеры погружений Купмана	326
Аналитическое разложение собственных функций в ряд	329
История и недавние достижения.....	331
7.5. Управляемый данными анализ Купмана	332
Расширенный DMD	332
Аппроксимация собственных функций Купмана на основе данных	334
Управляемый данными анализ Купмана и запаздывающие координаты	336
Нейронные сети для погружений Купмана.....	340
Рекомендуемая литература	342

Глава 8. Теория линейного управления

344

Типы управления	345
8.1. Управление с замкнутым контуром обратной связи.....	346
Примеры преимуществ управления с обратной связью.....	348
8.2. Линейные стационарные системы	351
Линеаризация нелинейной динамики	351

Неуправляемая линейная система	352
Управляемая линейная система	354
Системы с дискретным временем	355
Пример: обратный маятник	356
8.3. Управляемость и наблюдаемость	357
Управляемость	357
Наблюдаемость	359
Критерий управляемости РВН	360
Теорема Кэли–Гамильтона и достижимость	361
Грамианы и степень управляемости и наблюдаемости	362
Стабилизируемость и распознаваемость	364
8.4. Оптимальное управление полным состоянием: линейно-квадратичный регулятор (ЛКР)	364
Вывод уравнения Риккати оптимального управления	366
8.5. Оптимальное оценивание полного состояния: фильтр Калмана	369
8.6. Оптимальное управление с использованием датчиков: линейно-квадратичное гауссово управление (ЛКГ)	372
8.7. Практический пример: обратный маятник на тележке	374
Управление маятником на тележке с обратной связью	376
Оценка полного состояния системы маятник–тележка	379
Управление с обратной связью системой маятник–тележка с использованием датчиков	382
8.8. Робастное управление и методы анализа в частотной области	384
Методы в частотной области	384
Качество управления и передаточная функция контура: чувствительность и дополнительная чувствительность	389
Обращение динамики	392
Робастное управление	393
Рекомендуемая литература	396

Глава 9. Сбалансированные модели, пригодные

для управления	397
9.1. Упрощение модели и идентификация системы	397
9.2. Сбалансированное упрощение модели	399
Цель упрощения модели	399
Замена переменных в системах управления	401
Балансирующие преобразования	403
Сбалансирование усечения	407
Вычисление сбалансированных реализаций	408
Пример сбалансированного упрощения модели	413
9.3. Идентификация системы	415
Алгоритм реализации собственной системы	416
Идентификация наблюдателей с помощью фильтра Калмана	419
Комбинация ERA и OKID	423
Рекомендуемая литература	425

Глава 10. Управление на основе данных	426
10.1. Идентификация нелинейной системы для управления.....	427
DMD с управлением.....	428
Нелинейное управление с помощью оператора Купмана.....	430
SINDy с управлением.....	432
Пример управления на основе прогнозирующих моделей (MPC).....	432
10.2. Управление с машинным обучением.....	436
Обучение с подкреплением.....	438
Управление с итеративным обучением.....	439
Генетические алгоритмы.....	439
Генетическое программирование.....	441
Пример: применение генетического алгоритма для настройки ПИД-регулятора.....	443
10.3. Адаптивное управление с поиском экстремума.....	448
Простой пример управления с поиском экстремума.....	452
Пример управления с поиском экстремума в сложной ситуации.....	455
Приложения управления с поиском экстремума.....	456
Рекомендуемая литература.....	458
Часть IV. МОДЕЛИ ПОНИЖЕННОГО ПОРЯДКА	460
Глава 11. Модели пониженного порядка (ROM)	461
11.1. POD для дифференциальных уравнений в частных производных.....	462
Разложение по модам Фурье.....	465
Специальные функции и теория Штурма–Лиувилля.....	466
Понижение размерности.....	467
11.2. Элементы оптимального базиса: собственное ортогональное разложение.....	468
Проекция Галеркина на POD-моды.....	470
Пример: гармонический осциллятор.....	471
11.3. POD и динамика солитонов.....	475
Упрощение солитона ($N = 1$).....	477
Упрощение солитона ($N = 2$).....	479
11.4. POD в непрерывной формулировке.....	480
Квадратурные правила для R: правило трапеций.....	482
Квадратурные правила более высокого порядка.....	483
POD-моды и квадратурные формулы.....	485
11.5. POD с симметриями: повороты и сдвиги.....	486
Сдвиг: распространение волн.....	486
Поворот: спиральные волны.....	488
Рекомендуемая литература.....	492
Глава 12. Интерполяция для ROM	494
12.1. Неполное POD.....	494
Разреженные измерения и реконструкция.....	496

Моды гармонического осциллятора	497
12.2. Ошибка и сходимость неполного POD	501
Случайная выборка и сходимость	501
Неполные измерения и качество реконструкции.....	503
12.3. Неполные измерения: минимизация числа обусловленности	504
Замены числа обусловленности.....	510
12.4. Неполные измерения: максимальная дисперсия.....	512
12.5. POD и дискретный эмпирический метод интерполяции (DEIM) ...	517
POD и DEIM	518
DEIM	519
12.6. Реализация алгоритма DEIM.....	521
Алгоритм QDEIM	523
12.7. Машинное обучение ROM.....	524
Выбор POD-моды	525
Пример: обтекание цилиндра	527
Рекомендуемая литература	529
Глоссарий.....	531
Список литературы.....	538
Предметный указатель	539

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.

Скачивание исходного кода примеров

Скачать файлы с дополнительной информацией для книг издательства «ДМК Пресс» можно на сайте www.dmkpress.com на странице с описанием соответствующей книги.

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, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательства «ДМК Пресс» и Springer очень серьезно относятся 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 электронной почты dmkpress@gmail.com.

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

Об авторах

Стивен Л. Брантон – доцент факультета общего машиностроения в Вашингтонском университете. Также является внештатным сотрудником отделения прикладной математики и науки о данных Института eScience. Область его научных интересов охватывает применение науки о данных и машинного обучения к динамическим системам и управлению в области гидрогазодинамики, биомеханики, оптики, энергетических систем и производства. Является автором двух учебников, лауреатом премии армии и ВВС для молодых ученых, получил право преподавания в инженерном колледже Вашингтонского университета и удостоен премии для молодых преподавателей.

Дж. Натан Куц – профессор прикладной математики Вашингтонского университета, был деканом факультета до 2015 года. Также является внештатным профессором отделения электротехники и физики и старшим научным сотрудником отделения науки о данных в Институте eScience. Область научных интересов охватывает сложные системы и анализ данных, конкретно применение методов машинного обучения и динамических систем и управление в разнообразных приложениях. Автор двух учебников, лауреат премии Боинг за отличное преподавание прикладной математики, а также премии CAREER Национального научного фонда США.

Предисловие

Эта книга посвящена растущей области знаний на пересечении методов обработки больших данных, прикладной оптимизации и классических дисциплин инженерной математики и математической физики. Мы готовили данный материал на протяжении ряда лет, в основном для лекций, читаемых студентам старших курсов и аспирантам технических и физических факультетов.

Обычно такие студенты имеют подготовку в области линейной алгебры, дифференциальных уравнений и научных расчетов, а инженеры также знакомы с теорией управления и (или) дифференциальными уравнениями в частных производных. Однако в большинстве учебных программ научно-технических вузов методы обработки данных и (или) оптимизации освещаются слабо или не включены вовсе. С другой стороны, студенты, обучающиеся по специальностям «информатика» и «статистика», плохо знакомы с динамическими системами и теорией управления. Нашей целью было написать введение в прикладную науку о данных для обеих групп. Включенные в книгу методы отбирались по трем критериям: (1) релевантность, (2) простота и (3) общность. Мы стремились представить широкий круг тем от вводного материала до методов, реально применяемых в исследованиях.

Открытие на основе анализа данных революционизировало наши подходы к моделированию, прогнозированию поведения и управлению сложными системами. Самые насущные научно-технические задачи нашего времени не поддаются эмпирическим моделям и выводам, основанным на первопринципах. Все чаще исследователи обращаются к подходам на основе анализа данных при изучении широкого спектра сложных систем, как то: турбулентность, науки о мозге, климатология, эпидемиология, финансы, робототехника, автономные системы. Такие системы обычно являются нелинейными, динамическими, многомасштабными в пространстве и во времени, многомерными и имеют доминирующие паттерны, которые необходимо охарактеризовать и смоделировать, чтобы в конечном итоге обеспечить сбор данных, прогнозирование, оценку и управление. Благодаря современным математическим методам вкупе с невиданной ранее доступностью данных и располагаемыми вычислительными ресурсами мы теперь можем подступить к неприступным до недавнего времени проблемам. Упомянем лишь малую толику новых методов: надежное восстановление изображения по разреженным и зашумленным измерениям случайных пикселей, управление турбулентностью с помощью машинного обучения, оптимальное размещение датчиков и приводов, идентификация допускающих интерпретацию нелинейных динамических систем на основе одних лишь данных и модели пониженного порядка, позволяющие ускорить изучение и оптимизацию систем со сложной многомасштабной физикой.

Движущим началом современной науки о данных является доступность больших и постоянно увеличивающихся объемов данных вследствие заме-

чательных инноваций в области разработки дешевых датчиков, возросших на порядки вычислительных мощностей и практически неограниченной емкости устройств хранения и скорости передачи. Такое изобилие данных открывает перед учеными и инженерами во всех областях новые возможности для изобретений на основе анализа данных; часто в этой связи говорят о четвертой парадигме научного открытия [245]. Эта четвертая парадигма представляет собой естественную кульминацию первых трех: эмпирического эксперимента, аналитического вывода и вычислительного исследования. Интеграция всех трех методик создает новаторскую платформу для новых открытий на основе данных. Этот процесс научного открытия не нов и по сути дела повторяет усилия титанов научной революции: Иоганна Кеплера (1571–1630) и сэра Исаака Ньютона (1642–1727). Оба сыграли ключевую роль в разработке теоретических принципов небесной механики на базе сочетания эмпирических подходов, основанных на анализе данных, и аналитических вычислений. Наука о данных не заменяет математическую физику и технику, но дополняет ее с учетом достижений XXI века, что больше напоминает возрождение, нежели революцию.

Наука о данных сама по себе не нова, ее предложил больше 50 лет назад Джон Тьюки, предвидевший появление науки, в центре внимания которой будет обучение на данных, или *анализ данных* [152]. С тех пор в науке о данных преобладают два разных подхода [78]. Сообщество *машинного обучения* состоит в основном из специалистов по информатике и интересуется в первую очередь разработкой быстрых, масштабируемых и качественных алгоритмов прогнозирования. Сообщество же *статистического обучения*, которое вовсе необязательно во всем противопоставлять первому, больше сосредоточено на факультетах математической статистики и занимается выводом интерпретируемых моделей. Обе методологии могут похвастаться значительными успехами и закладывают математические и вычислительные основания методов науки о данных. Целью ученых и инженеров должно стать использование этих методов для выведения из результатов наблюдений и обсчета моделей (чаще нелинейных), которые правильно улавливают динамику системы и количественно и качественно обобщаются на ненаблюдавшиеся области фазового, параметрического или прикладного пространства. А в этой книге нашей целью будет применение статистических методов и методов машинного обучения к решению технических задач.

РАССМАТРИВАЕМЫЕ ВОПРОСЫ

В книге обсуждается целый ряд ключевых тем. Во-первых, во многих сложных системах присутствуют доминирующие *низкоразмерные паттерны* данных, несмотря на быстрое увеличение разрешающей способности измерений и вычислений. Базовая структура открывает возможность эффективного размещения датчиков и компактного представления для моделирования и управления. Выделение паттернов тесно связано со второй темой: отысканием *преобразований координат*, позволяющих упростить систему.

Действительно, богатая история математической физики вращается вокруг преобразований координат (например, спектральные разложения, преобразование Фурье, обобщенные функции и т. д.), хотя эти методы в большинстве своем были ограничены простой идеализированной геометрией и линейной динамикой. Умение выводить преобразования *на основе данных* открывает возможность обобщить их на новые задачи с более сложной геометрией и граничными условиями. На протяжении всей книги мы будем интересоваться *динамическими системами и управлением*, т. е. применением методов, основанных на анализе данных, к моделированию и управлению систем, изменяющихся во времени. Красной нитью проходит тема *управляемой данными прикладной оптимизации*, поскольку едва ли не каждый рассматриваемый вопрос так или иначе связан с оптимизацией (например, нахождение *оптимальных* низкоразмерных паттернов, *оптимальное* расположение датчиков, *оптимизация* в машинном обучении, *оптимальное* управление и т. д.). И еще одна, даже более фундаментальная тема – большинство данных организовано в массивы для анализа, а широкое развитие численных инструментов линейной алгебры, начиная с 1960-х годов, лежит в основе математических методов матричных разложений и стратегий решения, встречающихся в этой книге.

БЛАГОДАРНОСТИ

Мы в долгу перед многими талантливыми студентами, сотрудниками и коллегами, которые делились ценными замечаниями и предложениями и оказывали нам поддержку. Особенно мы благодарны Джошуа Проктору (Joshua Proctor), который стоял у истоков этой книги и помогал при планировании ее структуры и организации. Мы также извлекли много полезного из бесед с Бингом Брантоном (Bing Brunton), Игорем Мезичем (Igor Mezić), Берндом Ноаком (Bernd Noack) и Сэмом Тайрой (Sam Taira). Эта книга была бы невозможна без помощи наших сотрудников и коллег, исследования которых отражены в тексте.

На протяжении работы над книгой и чтения соответствующих курсов мы получали чрезвычайно ценные отзывы и комментарии от наших замечательных студентов и постдоков: Трэвиса Эшкама (Travis Askham), Майкла Ау-Юнга (Michael Au-Yeung), Цзе Бая (Zhe Bai), Идо Брайта (Ido Bright), Кэтлин Чемпион (Kathleen Champion), Эмили Кларк (Emily Clark), Чарльза Делаханта (Charles Delahunt), Даниэля Дылевски (Daniel Dylewski), Бена Эрричсона (Ben Erichson), Чарли Фислера (Charlie Fiesler), Синь Фу (Xing Fu), Чена Гонга (Chen Gong), Тарена Гормана (Taren Gorman), Джейкоба Гросека (Jacob Grosek), Сета Хирша (Seth Hirsh), Микала Джонсона (Mikala Johnson), Юрики Кайзер (Eurika Kaiser), Мейсона Кема (Mason Kamb), Джеймса Кьюнерта (James Kunert), Бетани Луш (Bethany Lusch), Педро Майа (Pedro Maia), Критики Манохара (Krithika Manohar), Найла Мэнгана (Niall Mangan), Арианы Мендибль (Ariana Mendible), Томаса Морена (Thomas Mohren), Меган Моррисон (Megan Morrison), Маркуса Куэйда (Markus Quade), Сэма Руди (Sam Rudy), Сюзанны Саргсян

(Susanna Sargsyan), Изабель Шерл (Isabel Scherl), Эли Шлизермана (Eli Shlizerman), Джорджа Степанянца (George Stepaniants), Бена Строма (Ben Strom), Чань Суна (Chang Sun), Роя Тэйлора (Roy Taylor), Мегханы Велагар (Meghana Velagar), Джейка Вехолта (Jake Weholt) и Мэтта Уильямса (Matt Williams). Наши студенты подвигли нас на написание этой книги, благодаря им мы каждый день приходим на работу с радостью и волнением.

Мы также благодарны руководителю издательской группы в Cambridge University Press Лоран Коулз (Lauren Cowles), на которую могли положиться на протяжении всего процесса работы.

ОНЛАЙНОВЫЕ МАТЕРИАЛЫ

Мы с самого начала предполагали, что к книге будут прилагаться обширные дополнительные материалы: код, данные, видео, домашние задания и рекомендуемые способы построения курса. Все эти материалы можно найти на сайте databookuw.com.

Код в сети полнее, чем в книге, в частности включен код генерации рисунков, пригодных для публикации. Визуализация данных была поставлена на первое место среди методов науки о данных в опросе «Состояние науки о данных и машинного обучения», проведенном на Kaggle 2017. Поэтому мы настоятельно рекомендуем читателям скачать код с сайта и в полной мере воспользоваться командами построения графиков и диаграмм.

Мы также записали и разместили на YouTube лекции по большинству тем, включенных в книгу. Есть дополнительные видео для студентов, желающих восполнить пробелы в подготовке по научным расчетам и основам прикладной математики. Этот текст задуман одновременно как справочное пособие и источник материалов к нескольким курсам, рассчитанным на студентов разного уровня. Большинство глав самостоятельны, на их основе можно разработать *курсы молодого бойца*, рассчитанные примерно на 10 часов каждый.

КАК ПОЛЬЗОВАТЬСЯ ЭТОЙ КНИГОЙ

Книга рассчитана на начинающих аспирантов и продвинутых студентов старших курсов научных и технических факультетов. Поэтому методы машинного обучения излагаются с азов, но при этом мы предполагаем, что студенты умеют моделировать физические системы с помощью дифференциальных уравнений и решать их с помощью таких программ, как **ode45**. Рассматриваются как начальные вопросы, так и актуальные исследовательские методы. Наша цель – представить цельный взгляд и математический инструментарий для решения научно-технических задач. Но книга может быть также полезна студентам, изучающим информатику и статистику, которые зачастую мало знают о динамических системах и теории управления. На основе представленного материала можно разработать несколько курсов,

программы некоторых из них имеются на сайте книги и включают домашние задания, наборы данных и код.

Прежде всего мы хотели, чтобы книга была интересной, чтобы она вдохновляла, открывала глаза и вооружала знаниями молодых ученых и инженеров. Мы пытались по возможности не слишком усложнять, не жертвуя при этом глубиной и широтой охвата, без которых не может быть никакой исследовательской работы. Многие главы можно было бы развернуть в целые книги, и такие книги есть. Однако мы также стремились к полноте в той мере, в какой этого можно ожидать от книги, посвященной столь обширной и быстро развивающейся области. Мы надеемся, что книга придется вам по вкусу, что вы овладеете всеми описанными в ней методами и измените мир с помощью прикладной науки о данных!

Общеупотребительные методы оптимизации, уравнения, символы и акронимы

НАИБОЛЕЕ РАСПРОСТРАНЕННЫЕ СТРАТЕГИИ ОПТИМИЗАЦИИ

Метод наименьших квадратов (обсуждается в главах 1 и 4) минимизирует сумму квадратов разностей (ошибок) между фактическими данными и предсказаниями модели. В случае линейного метода наименьших квадратов, когда данные аппроксимируются линейной функцией, имеется решение в замкнутой форме, которое можно найти, приравняв к нулю производную ошибки по каждому неизвестному. Этот подход широко используется в технике и прикладных науках для аппроксимации полиномиальными функциями. Применение нелинейного метода наименьших квадратов обычно требует итеративного уточнения путем аппроксимации нелинейного решения линейным на каждой итерации.

Градиентный спуск (обсуждается в главах 4 и 6) – основной метод выпуклой оптимизации в многомерных системах. Для минимизации ошибки вычисляется градиент аппроксимирующей функции. Решение обновляется итеративно путем *спуска с горы* в пространстве решений. Одномерным вариантом градиентного спуска является метод Ньютона–Рафсона. В многомерном пространстве метод часто находит только локальный минимум. Важнейшими инновациями в приложениях больших данных являются стохастический градиентный спуск и алгоритм обратного распространения, благодаря чему оптимизация сводится к самому вычислению градиента.

Чередующийся градиентный спуск (Alternating Descent Method – ADM) (обсуждается в главе 4) позволяет избежать вычисления градиента за счет того, что на каждом шаге производится оптимизация по одной неизвестной. Таким образом, все неизвестные переменные считаются постоянными, за исключением одной, по которой производится линейный поиск (выпуклая оптимизация). Эта переменная обновляется, после чего фиксируется, и то же самое повторяется для другой переменной. На одном шаге итерации пере-

бираются все неизвестные, а сами итерации продолжают до тех пор, пока не будет достигнута желаемая точность.

Расширенный метод Лагранжа (Augmented Lagrange Method – ALM) (обсуждается в главах 3 и 8) – класс алгоритмов для решения задач условной оптимизации. Они похожи на методы штрафования тем, что заменяют задачу оптимизации с ограничениями последовательностью задач без ограничений и прибавляют к целевой функции штрафной член, который играет роль множителя Лагранжа. Расширенный метод Лагранжа – не то же самое, что метод множителей Лагранжа.

Линейное программирование и симплекс-метод – безотказные алгоритмы выпуклой оптимизации. В линейном программировании целевая функция линейно зависит от неизвестных, а ограничениями являются линейные равенства и неравенства. Вычислив область допустимых решений – выпуклый политоп, – алгоритм линейного программирования находит в полиэдре точку, в которой функция принимает наименьшее (или наибольшее) значение, если таковая существует. Симплекс-метод – это конкретная итеративная процедура линейного программирования, которая по заданному опорному допустимому решению пытается найти другое опорное решение, для которого целевая функция принимает меньшее значение, и тем самым производит оптимизацию.

НАИБОЛЕЕ УПОТРЕБИТЕЛЬНЫЕ УРАВНЕНИЯ И СИМВОЛЫ

Линейная алгебра

Линейная система уравнений

$$Ax = b. \quad (0.1)$$

Матрица $A \in \mathbb{R}^{p \times n}$ и вектор $b \in \mathbb{R}^p$ обычно известны, а вектор $x \in \mathbb{R}^n$ неизвестен.

Уравнение для собственных значений

$$AT = T\Lambda. \quad (0.2)$$

Столбец ξ_k матрицы T является собственным вектором матрицы $A \in \mathbb{C}^{n \times n}$, соответствующим собственному значению λ_k : $A\xi_k = \lambda_k\xi_k$. Матрица Λ – диагональная матрица, содержащая эти собственные значения, в простейшем случае все n собственных значений различны.

Замена координат

$$x = \Psi a. \quad (0.3)$$

Вектор $x \in \mathbb{R}^n$ можно записать как $a \in \mathbb{R}^n$ в системе координат, определенной столбцами матрицы $\Psi \in \mathbb{R}^{n \times n}$.

Уравнение измерений

$$\mathbf{y} = \mathbf{C}\mathbf{x}. \tag{0.4}$$

Вектор $\mathbf{y} \in \mathbb{R}^p$ является измерением состояния $\mathbf{x} \in \mathbb{R}^n$ в результате применения матрицы измерений $\mathbf{C} \in \mathbb{R}^{p \times n}$.

Сингулярное разложение

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^*. \tag{0.5}$$

Матрицу $\mathbf{X} \in \mathbb{C}^{n \times m}$ можно разложить в произведение трех матриц $\mathbf{U} \in \mathbb{C}^{n \times n}$, $\mathbf{\Sigma} \in \mathbb{C}^{n \times m}$ и $\mathbf{V} \in \mathbb{C}^{m \times m}$. Матрицы \mathbf{U} и \mathbf{V} унитарные, т. е. $\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}^{n \times n}$ и $\mathbf{V}\mathbf{V}^* = \mathbf{V}^*\mathbf{V} = \mathbf{I}^{m \times m}$, где $*$ обозначает операцию комплексного сопряжения и транспонирования. Столбцы \mathbf{U} (соответственно \mathbf{V}) ортогональны и называются левыми (соответственно правыми) *сингулярными векторами*. На главной диагонали диагональной матрицы $\mathbf{\Sigma}$ находятся убывающие неотрицательные элементы, называемые *сингулярными значениями*.

Часто \mathbf{X} аппроксимируется матрицей низкого ранга $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^*$, где $\tilde{\mathbf{U}}$ и $\tilde{\mathbf{V}}$ содержат первые $r \ll n$ столбцов \mathbf{U} и \mathbf{V} соответственно, а $\tilde{\mathbf{\Sigma}}$ – левый верхний блок $\mathbf{\Sigma}$ размера $r \times r$. В контексте пространственных мод, моделей пониженного порядка и размещения датчиков матрица $\tilde{\mathbf{U}}$ часто обозначается буквой $\mathbf{\Psi}$.

Регрессия и оптимизация

Оптимизация переопределенных и недоопределенных линейных систем

$$\operatorname{argmin}_{\mathbf{x}} (\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 + \lambda g(\mathbf{x})) \text{ или} \tag{0.6a}$$

$$\operatorname{argmin}_{\mathbf{x}} g(\mathbf{x}) \text{ при условии } \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \varepsilon. \tag{0.6b}$$

Здесь $g(\mathbf{x})$ – штраф регрессии (со штрафным параметром λ для переопределенных систем). Для переопределенных и недоопределенных систем линейных уравнений $\mathbf{A}\mathbf{x} = \mathbf{b}$, когда решений либо не существует, либо бесконечно много, для нахождения решения нужно задать ограничение или штраф; эта процедура называется *регуляризацией*.

Оптимизация переопределенных и недоопределенных линейных систем

$$\operatorname{argmin}_{\mathbf{x}} (f(\mathbf{A}, \mathbf{x}, \mathbf{b}) + \lambda g(\mathbf{x})) \text{ или} \tag{0.7a}$$

$$\operatorname{argmin}_{\mathbf{x}} g(\mathbf{x}) \text{ при условии } f(\mathbf{A}, \mathbf{x}, \mathbf{b}) \leq \varepsilon. \tag{0.7b}$$

Это обобщение линейной системы на нелинейную систему $f(\cdot)$ с регуляризацией $g(\cdot)$. Такие переопределенные и недоопределенные системы часто решаются методами градиентного спуска.

Композиционная оптимизация для нейронных сетей

$$\operatorname{argmin}_{\mathbf{A}_j} (f_M(\mathbf{A}_M, \dots, f_2(\mathbf{A}_2, (f_1(\mathbf{A}_1, \mathbf{x})) \dots)) + \lambda g(\mathbf{A}_j)). \tag{0.8}$$

Здесь \mathbf{A}_k – матрицы весов связей между k -м и $(k + 1)$ -м слоями нейронной сети. Обычно это сильно недоопределенная система, которая регуляризи-

руется прибавлением $g(\mathbf{A}_j)$. Композиция и регуляризация весьма важны как для порождения выразительных представлений данных, так и для предотвращения переобучения.

Динамические системы и системы пониженного порядка

*Нелинейное обыкновенное дифференциальное уравнение
(динамическая система)*

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), t; \boldsymbol{\beta}). \quad (0.9)$$

Вектор $\mathbf{x}(t) \in \mathbb{R}^n$ описывает состояние системы, изменяющейся во времени t , $\boldsymbol{\beta}$ – вектор параметров, а \mathbf{f} – векторное поле. В общем случае \mathbf{f} – липшицева функция, что гарантирует существование и единственность решения.

Система с линейной зависимостью выхода от входа

$$\frac{d}{dt} \mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (0.10a)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}. \quad (0.10b)$$

Состояние системы представлено вектором $\mathbf{x} \in \mathbb{R}^n$, входы (приводы) – вектором $\mathbf{u} \in \mathbb{R}^q$, а выходы (датчики) – вектором $\mathbf{y} \in \mathbb{R}^p$. Матрицы \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{D} определяют динамику, управляющее воздействие, стратегию работы датчиков и эффект сквозного управления соответственно.

*Нелинейное отображение
(динамические системы с дискретным временем)*

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k). \quad (0.11)$$

Состояние системы на k -й итерации представлено вектором $\mathbf{x}_k \in \mathbb{R}^n$, а \mathbf{F} – потенциально нелинейное отображение. Часто это отображение описывает продвижение итераций во времени, т. е. $\mathbf{x}_k = \mathbf{x}(k\Delta t)$; в таком случае потоковое отображение обозначается $\mathbf{F}_{\Delta t}$.

Операторное уравнение Купмана (с дискретным временем)

$$\mathcal{K}g = g \circ \mathbf{F}_t \Rightarrow \mathcal{K}_t \varphi = \lambda \varphi. \quad (0.12)$$

Линейный оператор Купмана \mathcal{K}_t экстраполирует функции измерения состояния $g(\mathbf{x})$ с помощью потока \mathbf{F}_t . Собственные значения и собственные векторы \mathcal{K}_t обозначаются λ и $\varphi(\mathbf{x})$ соответственно. Оператор \mathcal{K}_t применяется к гильбертову пространству измерений.

Нелинейные дифференциальные уравнения в частных производных (УрЧП)

$$\mathbf{u}_t = \mathbf{N}(\mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}, \dots, x, t; \boldsymbol{\beta}). \quad (0.13)$$

Состояние УрЧП описывается вектором \mathbf{u} , \mathbf{N} – нелинейный оператор эволюции, нижние индексы обозначают взятие частных производных, а x и t –

пространственная и временная переменные соответственно. УрЧП параметризуется значениями, собранными в векторе β . Состояние УрЧП \mathbf{u} может быть непрерывной функцией $u(x, t)$, а может быть дискретизировано в нескольких точках пространства, $\mathbf{u}(t) = [u(x_1, t) \ u(x_2, t) \ \dots \ u(x_n, t)]^T \in \mathbb{R}^n$.

Разложение Галеркина

Непрерывное разложение Галеркина имеет вид:

$$u(x, t) \approx \sum_{k=1}^r a_k(t) \psi_k(x). \tag{0.14}$$

Функции $a_k(t)$ – коэффициенты, отражающие временную динамику, а $\psi_k(x)$ – пространственные моды. Для многомерного дискретизированного состояния разложение Галеркина принимает вид $\mathbf{u}(t) \approx \sum_{k=1}^r a_k(t) \Psi_k$. Пространственные моды $\Psi_k \in \mathbb{R}^n$ могут быть столбцами матрицы $\Psi = \tilde{\mathbf{U}}$.

СПИСОК ОБОЗНАЧЕНИЙ

Размерности

- K количество ненулевых элементов K -разреженного вектора \mathbf{s}
- m количество снимков данных (т. е. столбцов \mathbf{X})
- n размерность состояния $\mathbf{x} \in \mathbb{R}^n$
- p размерность измерения, или выходной переменной $\mathbf{y} \in \mathbb{R}^p$
- q размерность выходной переменной $\mathbf{u} \in \mathbb{R}^q$
- r ранг усеченного сингулярного разложения или иной низкоранговой аппроксимации

Скаляры

- s частота в лапласовой области
- t время
- δ скорость обучения в методе градиентного спуска
- Δt временной шаг
- x пространственная переменная
- Δx пространственный шаг
- σ сингулярное значение
- λ собственное значение
- λ параметр разреженности при разреженной оптимизации (раздел 7.3)
- λ множитель Лагранжа (разделы 3.7, 8.4 и 11.4)
- τ порог

Векторы

- \mathbf{a} вектор амплитуд мод \mathbf{x} в базисе Ψ , $\mathbf{a} \in \mathbb{R}^r$
- \mathbf{b} вектор измерений в линейной системе $\mathbf{Ax} = \mathbf{b}$
- \mathbf{b} вектор амплитуд мод в разложении по динамическим модам (раздел 7.2)
- \mathbf{Q} вектор, содержащий функцию потенциала в алгоритме PDE-FIND

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru