

УДК 004.85
ББК 32.971.3
Г52

Гласснер Э.

Г52 Глубокое обучение без математики. Т. 1: Основы / пер. с англ. В. А. Яроцкого. – М.: ДМК Пресс, 2019. – 584 с.: ил.

ISBN 978-5-97060-701-5

Эта книга не похожа на большинство других учебников и руководств по глубокому обучению – в ней нет ни детального алгоритмического анализа, сопровождаемого обширной математикой, ни развернутых листингов программного кода. Автор выбрал золотую середину — благодаря дружелюбному подходу, сопровождаемому огромным количеством цветных иллюстраций, а также детальному и скрупулезному описанию он глубоко освещает основополагающие идеи и фундаментальные основы глубокого обучения и нейронных сетей.

Издание предназначено для всех читателей, кто хочет использовать глубокое обучение и нейронные сети в своей работе.

Продолжение книги будет посвящено практическому воплощению алгоритмов глубокого обучения и выйдет в издательстве «ДМК Пресс» летом 2019 года.

УДК 004.85
ББК 32.971.3

Original English language edition published by The Imaginary Institute, Seattle, WA. Copyright © 2018 by Andrew Glassner. Russian-language edition copyright © 2019 by DMK Press. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

ISBN 978-5-97060-701-5 (рус.)

Copyright © 2018 by Andrew Glassner
© Оформление, издание, перевод, ДМК Пресс, 2019

Содержание

ТОМ 1. ОСНОВЫ

| | |
|---|----|
| Предисловие. Добро пожаловать! | 12 |
|---|----|

| | |
|--|----|
| Глава 1. Введение в машинное обучение и глубокое обучение | 19 |
|--|----|

| | |
|---|----|
| 1.1. Зачем здесь эта глава | 19 |
| 1.1.1. Извлечение значащей информации из данных | 20 |
| 1.1.2. Экспертные системы | 22 |
| 1.2. Изучение маркированных данных | 23 |
| 1.2.1. Стратегия обучения | 25 |
| 1.2.2. Стратегия компьютерного обучения | 26 |
| 1.2.3. Обобщение | 28 |
| 1.2.4. Более внимательный взгляд на обучение | 30 |
| 1.3. Обучение с учителем | 31 |
| 1.3.1. Классификация | 32 |
| 1.3.2. Регрессия | 33 |
| 1.4. Обучение без учителя | 34 |
| 1.4.1. Кластеризация | 35 |
| 1.4.2. Подавление шума | 35 |
| 1.4.3. Понижение размерности | 36 |
| 1.5. Генераторы | 40 |
| 1.6. Обучение с подкреплением..... | 41 |
| 1.7. Глубокое обучение | 43 |
| 1.8. Что последует дальше | 47 |
| Справочные материалы | 47 |
| Заимствованные рисунки | 48 |

| | |
|--|----|
| Глава 2. Хаотичность и базовая статистика | 49 |
|--|----|

| | |
|---|----|
| 2.1. Зачем здесь эта глава | 49 |
| 2.2. Случайные переменные | 50 |
| 2.2.1. Случайные числа на практике | 55 |
| 2.3. Некоторые общепотребительные распределения | 57 |
| 2.3.1. Равномерное распределение | 57 |
| 2.3.2. Нормальное распределение..... | 58 |
| 2.3.3. Распределение Бернулли | 62 |
| 2.3.4. Мультиномиальное распределение | 63 |
| 2.3.5. Математическое ожидание | 64 |
| 2.4. Зависимость | 64 |
| 2.4.1. Независимые, одинаково распределенные переменные | 65 |
| 2.5. Выбор и возврат | 65 |
| 2.5.1. Выбор с возвращением | 66 |
| 2.5.2. Выбор без возвращения | 66 |
| 2.5.3. Осуществление выбора | 67 |
| 2.6. Оценка ошибки обобщения. Бутстраппинг..... | 68 |
| 2.7. Многомерные пространства..... | 72 |
| 2.8. Ковариация и корреляция | 74 |
| 2.8.1. Ковариация | 74 |

| | |
|-----------------------------|----|
| 2.8.2. Корреляция | 76 |
| 2.9. Квартет Энскомба | 79 |
| Справочные материалы | 81 |

Глава 3. Вероятность..... 82

| | |
|---|-----|
| 3.1. Почему здесь эта глава | 82 |
| 3.2. Метание дротика | 83 |
| 3.3. Простая вероятность..... | 85 |
| 3.4. Условная вероятность | 85 |
| 3.5. Совместная вероятность..... | 89 |
| 3.6. Маргинальная вероятность | 92 |
| 3.7. Корректность измерений..... | 93 |
| 3.7.1. Классификация выборок | 93 |
| 3.7.2. Матрица неточностей | 96 |
| 3.7.3. Интерпретация матрицы неточностей | 98 |
| 3.7.4. Когда неправильная классификация приемлема | 100 |
| 3.7.5. Верность..... | 102 |
| 3.7.6. Точность..... | 103 |
| 3.7.7. Полнота | 104 |
| 3.7.8. О точности и полноте | 105 |
| 3.7.9. Другие критерии | 108 |
| 3.7.10. Совместное использование точности и полноты | 109 |
| 3.7.11. Мера f_1 | 112 |
| 3.8. Применение матрицы неточностей | 113 |
| Справочные материалы | 118 |

Глава 4. Правило Байеса..... 119

| | |
|--|-----|
| 4.1. Почему здесь эта глава | 119 |
| 4.2. Частотная и байесовская вероятности | 120 |
| 4.2.1. Частотный подход..... | 120 |
| 4.2.2. Байесовский подход | 120 |
| 4.2.3. Обсуждение..... | 121 |
| 4.3. Подбрасывание монеты..... | 122 |
| 4.4. Это несмещенная монета? | 122 |
| 4.4.1. Правило Байеса..... | 131 |
| 4.4.2. Замечания по поводу правила Байеса..... | 133 |
| 4.5. Поиски жизни..... | 135 |
| 4.6. Повторные применения правила Байеса | 139 |
| 4.6.1. Цикл постериор–приор..... | 139 |
| 4.6.2. Пример: какую монету мы имеем? | 141 |
| 4.7. Множественные гипотезы | 146 |
| Справочные материалы | 153 |

Глава 5. Кривые и поверхности..... 154

| | |
|-----------------------------------|-----|
| 5.1. Почему здесь эта глава | 154 |
| 5.2. Введение | 154 |
| 5.3. Производная..... | 156 |
| 5.4. Градиент | 164 |
| Справочные материалы | 169 |

| | |
|--|------------|
| Глава 6. Теория информации | 170 |
| 6.1. Почему здесь эта глава | 170 |
| 6.1.1. Информация: одно слово, два значения | 170 |
| 6.2. Удивление и контекст | 170 |
| 6.2.1. Удивление | 171 |
| 6.2.2. Контекст | 172 |
| 6.3. Бит как единица информации | 173 |
| 6.4. Измерение информации | 173 |
| 6.5. Размер события | 175 |
| 6.6. Адаптивные коды | 175 |
| 6.7. Энтропия | 181 |
| 6.8. Кросс-энтропия | 183 |
| 6.8.1. Два адаптивных кода | 184 |
| 6.8.2. Смещение кодов | 186 |
| 6.9. Расходимость | 188 |
| Справочные материалы | 190 |
| Глава 7. Классификация | 192 |
| 7.1. Почему здесь эта глава | 192 |
| 7.2. Двумерная классификация | 193 |
| 7.2.1. 2D-бинарная классификация | 193 |
| 7.3. 2D-многоклассовая классификация | 197 |
| 7.4. Бинарная многоклассовая классификация | 199 |
| 7.4.1. Один против остальных | 199 |
| 7.4.2. Один против одного | 201 |
| 7.5. Кластеризация | 204 |
| 7.6. Проклятие размерности | 207 |
| 7.6.1. Странности большой размерности | 214 |
| Справочные материалы | 219 |
| Глава 8. Обучение и тестирование | 221 |
| 8.1. Почему здесь эта глава | 221 |
| 8.2. Обучение | 222 |
| 8.2.1. Тестирование качества обучения | 223 |
| 8.3. Тестовые данные | 226 |
| 8.4. Поверочные данные | 229 |
| 8.5. Кросс-валидация | 232 |
| 8.5.1. k -кратная валидация | 234 |
| 8.6. Использование результатов тестирования | 237 |
| Справочные материалы | 237 |
| Заимствованные изображения | 238 |
| Глава 9. Избыточное обучение и недостаточное обучение | 239 |
| 9.1. Почему здесь эта глава | 239 |
| 9.2. Избыточность и недостаточность | 240 |
| 9.2.1. Избыточность | 240 |
| 9.2.2. Недостаточность | 241 |
| 9.3. Избыточные данные | 241 |
| 9.4. Преждевременный останов | 245 |
| 9.5. Регуляризация | 247 |

| | |
|--|------------|
| 9.6. Смещение и дисперсия..... | 248 |
| 9.6.1. Аппроксимация базовых данных | 249 |
| 9.6.2. Большое смещение, маленькая дисперсия..... | 251 |
| 9.6.3. Маленькое смещение, большая дисперсия..... | 252 |
| 9.6.4. Сравнение кривых..... | 254 |
| 9.7. Аппроксимация по правилу Байеса | 256 |
| Справочные материалы | 262 |
| Глава 10. Нейроны | 263 |
| 10.1. Почему здесь эта глава..... | 263 |
| 10.2. Реальные нейроны | 263 |
| 10.3. Искусственные нейроны..... | 265 |
| 10.3.1. Перцептрон..... | 265 |
| 10.3.2. История перцептрона | 267 |
| 10.3.3. Современные искусственные нейроны | 268 |
| 10.4. Заключение..... | 272 |
| Справочные материалы | 273 |
| Глава 11. Обучение и мышление..... | 275 |
| 11.1. Почему здесь эта глава..... | 275 |
| 11.2. Ступени обучения | 275 |
| 11.2.1. Представление | 276 |
| 11.2.2. Оценка | 278 |
| 11.2.3. Оптимизация | 279 |
| 11.3. Дедукция и индукция | 280 |
| 11.4. Дедукция..... | 281 |
| 11.4.1. Ошибки категориальных силлогизмов | 285 |
| 11.5. Индукция | 287 |
| 11.5.1. Термины индукции в машинном обучении | 290 |
| 11.5.2. Ошибки индукции | 290 |
| 11.6. Объединенные умозаключения | 292 |
| 11.6.1. Шерлок Холмс – «мастер дедукции»..... | 293 |
| 11.7. Оперантное обуславливание | 294 |
| Справочные материалы | 296 |
| Глава 12. Подготовка данных | 298 |
| 12.1. Почему здесь эта глава..... | 298 |
| 12.2. Преобразование данных..... | 298 |
| 12.3. Типы данных | 301 |
| 12.3.1. Кодирование с использованием индивидуальных переменных | 302 |
| 12.4. Базовая очистка данных | 304 |
| 12.4.1. Очистка данных | 304 |
| 12.4.2. Очистка данных на практике..... | 305 |
| 12.5. Нормализация и стандартизация | 305 |
| 12.5.1. Нормализация..... | 306 |
| 12.5.2. Стандартизация | 307 |
| 12.5.3. Запоминание преобразований | 308 |
| 12.5.4. Типы преобразований..... | 309 |
| 12.6. Выбор характеристик..... | 310 |
| 12.7. Понижение размерности | 311 |
| 12.7.1. Анализ главных компонентов (PCA)..... | 311 |
| 12.7.2. Стандартизация и PCA для изображений..... | 317 |

| | |
|---|------------|
| 12.8. Преобразования | 324 |
| 12.9. Фрагментарная обработка | 329 |
| 12.9.1. Обработка по выборкам | 330 |
| 12.9.2. Обработка по характеристикам | 331 |
| 12.9.3. Обработка по элементам | 331 |
| 12.10. Преобразование с кросс-валидацией | 332 |
| Справочные материалы | 336 |
| Заимствованные изображения | 337 |
| Глава 13. Классификаторы | 338 |
| 13.1. Почему здесь эта глава | 338 |
| 13.2. Типы классификаторов | 339 |
| 13.3. Метод k -ближайших соседей (KNN) | 340 |
| 13.4. Метод опорных векторов (SVM) | 346 |
| 13.5. Деревья решений | 352 |
| 13.5.1. Построение деревьев | 357 |
| 13.5.2. Разделение узлов | 361 |
| 13.5.3. Контроль избыточного обучения | 363 |
| 13.6. Наивный Байес | 363 |
| 13.7. Обсуждение | 368 |
| Справочные материалы | 370 |
| Глава 14. Ансамбли | 371 |
| 14.1. Почему здесь эта глава | 371 |
| 14.2. Ансамбли | 371 |
| 14.3. Голосование | 372 |
| 14.4. Бутстрап-агрегация | 373 |
| 14.5. Случайные леса | 375 |
| 14.6. Рандомизированные деревья | 376 |
| 14.7. Бустинг | 377 |
| Справочные материалы | 384 |
| Глава 15. Библиотека Scikit-learn | 385 |
| 15.1. Почему здесь эта глава | 385 |
| 15.2. Введение | 386 |
| 15.3. Конвенции Python | 387 |
| 15.4. Оценщик | 390 |
| 15.4.1. Создание объекта | 391 |
| 15.4.2. Обучение с <code>fit()</code> | 392 |
| 15.4.3. Предсказание с <code>predict()</code> | 393 |
| 15.4.4. Функции <code>decision_function()</code> , <code>predict_proba()</code> | 395 |
| 15.5. Кластеризация | 395 |
| 15.6. Преобразователи | 398 |
| 15.6.1. Инверсные преобразования | 402 |
| 15.7. Уточнение данных | 405 |
| 15.8. Ансамбли | 407 |
| 15.9. Автоматизация | 410 |
| 15.9.1. Кросс-валидация | 410 |
| 15.9.2. Поиск гиперпараметров | 413 |
| 15.9.3. Поиск полным перебором по сетке | 416 |
| 15.9.4. Поиск случайным перебором по сетке | 424 |
| 15.9.5. Конвейер | 424 |

| | |
|--|-----|
| 15.9.6. Граница решения..... | 433 |
| 15.9.7. Конвейерные преобразования | 434 |
| 15.10. Наборы данных..... | 435 |
| 15.11. Утилиты..... | 437 |
| 15.12. Завершение..... | 440 |
| Справочные материалы | 440 |

Глава 16. Нейронные сети прямого распространения..... 442

| | |
|---|-----|
| 16.1. Почему здесь эта глава..... | 442 |
| 16.2. Графы нейронных сетей | 443 |
| 16.3. Синхронные и асинхронные потоки | 445 |
| 16.3.1. Графы в практике..... | 446 |
| 16.4. Инициализация весов..... | 447 |
| 16.4.1. Инициализация | 448 |
| Справочные материалы | 450 |

Глава 17. Функции активации..... 452

| | |
|--|-----|
| 17.1. Почему здесь эта глава..... | 452 |
| 17.2. Что делает функция активации | 452 |
| 17.2.1. Формы функций активации | 456 |
| 17.3. Основные функции активации..... | 456 |
| 17.3.1. Линейные функции..... | 456 |
| 17.3.2. Лестничная функция..... | 457 |
| 17.4. Ступенчатые функции | 458 |
| 17.5. Кусочно-линейные функции..... | 460 |
| 17.6. Гладкие функции | 463 |
| 17.7. Галерея функций активации | 467 |
| 17.8. Софтмакс | 468 |
| Справочные материалы | 470 |

Глава 18. Обратное распространение..... 471

| | |
|---|-----|
| 18.1. Почему здесь эта глава..... | 471 |
| 18.1.1. О тонкости настройки | 472 |
| 18.2. Очень медленный способ обучения | 473 |
| 18.2.1. Медленный способ обучения..... | 475 |
| 18.2.2. Более быстрый способ обучения | 477 |
| 18.3. Пока без функций активации..... | 479 |
| 18.4. Выходы нейронов и ошибка сети..... | 479 |
| 18.4.1. Пропорциональность изменения ошибок | 480 |
| 18.5. Маленькая нейронная сеть..... | 483 |
| 18.6. Шаг 1: дельты для выходных нейронов | 487 |
| 18.7. Шаг 2: использование дельт для изменения весов..... | 496 |
| 18.8. Шаг 3: дельты других нейронов | 499 |
| 18.9. Обратное распространение в действии..... | 504 |
| 18.10. Использование функций активации..... | 508 |
| 18.11. Скорость обучения | 514 |
| 18.11.1. Исследование скорости обучения..... | 515 |
| 18.12. Обсуждение | 523 |
| 18.12.1. Обратное распространение одной диаграммой..... | 523 |
| 18.12.2. Что обратное распространение не делает..... | 524 |
| 18.12.3. Что обратное распространение делает | 524 |

| | |
|--|-----|
| 18.12.4. Поддержка нейрона | 525 |
| 18.12.5. Мини-пакеты | 528 |
| 18.12.6. Параллельное обновление | 528 |
| 18.12.7. Чем привлекательно обратное распространение | 529 |
| 18.12.8. Обратное распространение не гарантировано | 529 |
| 18.12.9. Немного истории | 530 |
| 18.12.10. Погружение в математику | 530 |
| Справочные материалы | 532 |

Глава 19. Оптимизаторы..... 534

| | |
|--|-----|
| 19.1. Почему здесь эта глава..... | 534 |
| 19.2. Геометрия ошибки | 534 |
| 19.2.1. Минимумы, максимумы, плато и седла | 534 |
| 19.2.2. Двумерная кривая ошибки | 538 |
| 19.3. Настройка скорости обучения..... | 541 |
| 19.3.1. Обновления постоянного размера | 541 |
| 19.3.2. Изменение скорости обучения в процессе обучения..... | 548 |
| 19.3.3. План затуханий | 551 |
| 19.4. Стратегии обновления..... | 553 |
| 19.4.1. Пакетный градиентный спуск | 554 |
| 19.4.2. Стохастический градиентный спуск | 556 |
| 19.4.3. Мини-пакетный градиентный спуск | 558 |
| 19.5. Варианты градиентного спуска | 560 |
| 19.5.1. Метод импульса | 561 |
| 19.5.2. Импульс Нестерова..... | 567 |
| 19.5.3. Adagrad | 571 |
| 19.5.4. Adadelatа и RMSprop..... | 572 |
| 19.5.5. Adam | 574 |
| 19.6. Выбор оптимизатора | 575 |
| Справочные материалы | 576 |

Предметный указатель..... 578

ТОМ 2. ПРАКТИКА

Глава 20. Глубокое обучение

Глава 21. Нейронные сети сверток

Глава 22. Рекуррентные нейронные сети

Глава 23. Keras, часть1.

Глава 24. Keras, часть 2

Глава 25. Автокодировщики

Глава 26. Обучение с подкреплением

Глава 27. Порождающие состязательные сети

Глава 28. Применения для творчества

Глава 29. Наборы данных

Глава 30. Глоссарий