

Оглавление

Предисловие	17
Благодарности.....	19
О книге.....	20
Целевая аудитория.....	20
Структура издания и краткое содержание	21
О примерах кода	22
Об авторах	23
Иллюстрация на обложке	24
От издательства	25
О научном редакторе русскоязычного издания	25
Глава 1. Что такое глубокое обучение	26
1.1. Искусственный интеллект, машинное и глубокое обучение	27
1.2. Искусственный интеллект.....	27
1.3. Машинное обучение	28
1.4. Изучение правил и представлений данных	30
1.5. «Глубина» глубокого обучения.....	32
1.6. Принцип действия глубокого обучения в трех картинках.....	34
1.7. Отличительные черты глубокого обучения	36
1.8. Эпоха генеративного ИИ	37
1.9. Какой ступени развития достигло глубокое обучение	38
1.10. Не верьте поспешным обещаниям.....	39
1.11. Лето может смениться зимой.....	40
1.12. Перспективы ИИ.....	41
Глава 2. Математические основы нейронных сетей.....	43
2.1. Первое знакомство с нейронной сетью	44
2.2. Представление данных для нейронных сетей.....	48
2.2.1. Скаляры (тензоры нулевого ранга).....	49
2.2.2. Векторы (тензоры первого ранга)	49

2.2.3. Матрицы (тензоры второго ранга)	50
2.2.4. Тензоры третьего и более высоких рангов.....	50
2.2.5. Основные атрибуты	50
2.2.6. Манипулирование тензорами с помощью NumPy	52
2.2.7. Пакеты данных.....	52
2.2.8. Практические примеры тензоров с данными.....	53
2.3. Шестеренки нейронных сетей: операции с тензорами	56
2.3.1. Поэлементные операции	56
2.3.2. Расширение.....	58
2.3.3. Скалярное произведение тензоров.....	59
2.3.4. Изменение формы тензора	61
2.3.5. Геометрическая интерпретация операций с тензорами	62
2.3.6. Геометрическая интерпретация глубокого обучения	65
2.4. Механизм нейронных сетей: оптимизация на основе градиента	66
2.4.1. Что такое производная.....	68
2.4.2. Производная операций с тензорами — градиент	69
2.4.3. Стохастический градиентный спуск.....	71
2.4.4. Объединение производных: алгоритм обратного распространения ошибки.....	74
2.5. Оглядываясь на первый пример.....	78
2.5.1. Повторная реализация первого примера с нуля	80
2.5.2. Выполнение одного этапа обучения.....	82
2.5.3. Полный цикл обучения.....	84
2.5.4. Оценка модели	84
Итоги главы.....	85
Глава 3. Введение в TensorFlow, PyTorch, JAX и Keras.....	86
3.1. Краткая история фреймворков глубокого обучения	87
3.2. Связь между фреймворками.....	89
3.3. Введение в TensorFlow.....	90
3.3.1. Первые шаги с TensorFlow	91
3.3.2. Полный пример: линейный классификатор на TensorFlow	96
3.3.3. В чем уникальность работы с TensorFlow	100
3.4. Введение в PyTorch.....	101
3.4.1. Первые шаги с PyTorch.....	102
3.4.2. Полный пример: линейный классификатор на PyTorch	104
3.4.3. В чем уникальность работы с PyTorch	108
3.5. Введение в JAX.....	108
3.5.1. Первые шаги с JAX.....	109
3.5.2. Тензоры в JAX.....	109
3.5.3. Генерация случайных чисел в JAX	110
3.5.4. Полный пример: линейный классификатор на JAX.....	115
3.5.5. В чем уникальность работы с JAX?	116

8 Оглавление

3.6. Введение в Keras	117
3.6.1. Первые шаги с Keras	117
3.6.2. Слои — строительные блоки глубокого обучения	119
3.6.3. От слоев к моделям	122
3.6.4. Этап компиляции: настройка процесса обучения	124
3.6.5. Выбор функции потерь	125
3.6.6. Метод <code>fit()</code>	126
3.6.7. Оценка потерь и метрик на проверочных данных	127
3.6.8. Вывод: использование модели после обучения	128
Итоги главы	129
Глава 4. Классификация и регрессия	130
4.1. Классификация отзывов к фильмам: пример бинарной классификации	132
4.1.1. Набор данных IMDb	132
4.1.2. Подготовка данных	133
4.1.3. Конструирование модели	134
4.1.4. Проверка решения	137
4.1.5. Применение обученной сети для предсказаний на новых данных	141
4.1.6. Дальнейшие эксперименты	142
4.1.7. Подведение итогов	142
4.2. Классификация новостных лент: пример выделения нескольких классов	143
4.2.1. Набор данных Reuters	143
4.2.2. Подготовка данных	144
4.2.3. Конструирование модели	145
4.2.4. Проверка решения	146
4.2.5. Предсказания на новых данных	150
4.2.6. Другой способ обработки меток и потерь	150
4.2.7. Важность использования довольно больших промежуточных слоев	151
4.2.8. Дальнейшие эксперименты	151
4.2.9. Подведение итогов	152
4.3. Предсказание цен на дома: пример регрессии	152
4.3.1. Набор данных с ценами на жилье в Калифорнии	152
4.3.2. Подготовка данных	154
4.3.3. Конструирование модели	154
4.3.4. Оценка решения методом перекрестной проверки по K блокам	155
4.3.5. Предсказания на новых данных	159
4.3.6. Подведение итогов	160
Итоги главы	160
Глава 5. Основы машинного обучения	161
5.1. Обобщение — цель машинного обучения	161
5.1.1. Недообучение и переобучение	162
5.1.2. Природа обобщения в глубоком обучении	168

5.2. Оценка моделей машинного обучения.....	174
5.2.1. Обучающие, проверочные и контрольные наборы данных.....	174
5.2.2. Выбор базового уровня.....	178
5.2.3. Что важно помнить об оценке моделей.....	178
5.3. Улучшение качества обучения модели.....	179
5.3.1. Настройка основных параметров градиентного спуска.....	180
5.3.2. Использование более удачной архитектуры.....	181
5.3.3. Увеличение емкости модели.....	182
5.4. Улучшение способности к обобщению.....	185
5.4.1. Курирование набора данных.....	185
5.4.2. Конструирование признаков.....	186
5.4.3. Ранняя остановка.....	187
5.4.4. Регуляризация модели.....	188
Итоги главы.....	196
Глава 6. Обобщенный процесс машинного обучения.....	198
6.1. Определение задачи.....	199
6.1.1. Формулировка задачи.....	200
6.1.2. Сбор данных.....	202
6.1.3. Первичный анализ данных.....	206
6.1.4. Выбор меры успеха.....	206
6.2. Разработка модели.....	207
6.2.1. Подготовка данных.....	207
6.2.2. Выбор протокола оценки.....	209
6.2.3. Преодоление базового случая.....	209
6.2.4. Следующий шаг: разработка модели с переобучением.....	211
6.2.5. Регуляризация и настройка модели.....	212
6.3. Развертывание модели.....	213
6.3.1. Объяснение особенностей работы модели заинтересованным сторонам и обозначение границ ожидаемого.....	213
6.3.2. Предоставление доступа к модели.....	214
6.3.3. Мониторинг качества работы модели в процессе эксплуатации.....	217
6.3.4. Обслуживание модели.....	218
Итоги главы.....	219
Глава 7. Работа с Keras: глубокое погружение.....	220
7.1. Спектр рабочих процессов.....	221
7.2. Разные способы создания моделей Keras.....	222
7.2.1. Последовательная модель Sequential.....	222
7.2.2. Функциональный API.....	225
7.2.3. Создание производных от класса Model.....	232
7.2.4. Смешивание и согласование различных компонентов.....	234
7.2.5. Используйте правильный инструмент.....	235

7.3. Встроенные циклы обучения и оценки	236
7.3.1. Использование собственных метрик	237
7.3.2. Использование обратных вызовов.....	238
7.3.3. Разработка своего обратного вызова.....	241
7.3.4. Мониторинг и визуализация с помощью TensorBoard	242
7.4. Разработка своего цикла обучения и оценки	244
7.4.1. Обучение и прогнозирование	245
7.4.2. Создание своей функции шага обучения	246
7.4.3. Низкоуровневое использование метрик.....	251
7.4.4. Использование функции fit() с собственным циклом обучения	252
7.4.5. Обработка метрик в нестандартной функции train_step()	256
Итоги главы	259
Глава 8. Классификация изображений	261
8.1. Введение в сверточные нейронные сети	262
8.1.1. Операция свертывания	265
8.1.2. Выбор максимального значения из соседних	270
8.2. Обучение сверточной нейронной сети с нуля на небольшом наборе данных.....	273
8.2.1. Целесообразность глубокого обучения для решения задач с небольшими наборами данных	273
8.2.2. Загрузка данных	274
8.2.3. Конструирование сети.....	277
8.2.4. Предварительная обработка данных.....	279
8.2.5. Обогащение данных	283
8.3. Использование предварительно обученной модели	288
8.3.1. Выделение признаков.....	288
8.3.2. Дообучение предварительно обученной модели	296
Итоги главы	299
Глава 9. Архитектурные шаблоны сверточных сетей.....	300
9.1. Модульность, иерархия, многократное использование	301
9.2. Остаточные связи	304
9.3. Пакетная нормализация	308
9.4. Раздельная свертка по глубине.....	310
9.5. Собираем все вместе: мини-модель с архитектурой Xception	312
9.6. За пределами свертки. Трансформеры машинного зрения	314
Итоги главы	315
Глава 10. Интерпретация знаний, заключенных в сверточной нейронной сети	316
10.1. Визуализация промежуточных активаций.....	317
10.2. Визуализация фильтров сверточных нейронных сетей.....	323
10.2.1. Градиентный подъем в TensorFlow	326
10.2.2. Градиентный подъем в PyTorch	327

10.2.3. Градиентный подъем в JAX.....	327
10.2.4. Цикл визуализации фильтра.....	327
10.3. Визуализация тепловых карт активации класса.....	331
10.3.1. Получение градиентов для наиболее вероятного класса: версия с TensorFlow.....	334
10.3.2. Получение градиентов для наиболее вероятного класса: версия с PyTorch.....	334
10.3.3. Получение градиентов для наиболее вероятного класса: версия с JAX.....	335
10.3.4. Отображение тепловой карты активаций класса.....	336
10.4. Визуализация скрытого пространства сверточной нейронной сети.....	338
Итоги главы.....	339
Глава 11. Сегментация изображений.....	340
11.1. Задачи в сфере компьютерного зрения.....	340
11.1.1. Типы сегментации изображений.....	342
11.2. Обучение модели сегментации с нуля.....	343
11.2.1. Загрузка набора изображений для сегментации.....	343
11.2.2. Создание и обучение модели сегментации изображений.....	345
11.3. Использование предварительно обученной модели сегментации.....	349
11.3.1. Загрузка Segment Anything Model.....	350
11.3.2. Как работает Segment Anything.....	350
11.3.3. Подготовка контрольного изображения.....	352
11.3.4. Передача подсказки в виде целевой точки.....	354
11.3.5. Передача подсказки в виде целевой рамки.....	358
Итоги главы.....	359
Глава 12. Обнаружение объектов.....	360
12.1. Одно- и двухэтапные детекторы объектов.....	361
12.1.1. Двухэтапные детекторы R-CNN.....	362
12.1.2. Одноэтапные детекторы.....	363
12.2. Обучение модели YOLO с нуля.....	363
12.2.1. Загрузка набора данных COCO.....	364
12.2.2. Создание модели YOLO.....	367
12.2.3. Подготовка данных COCO для обучения модели YOLO.....	371
12.2.4. Обучение модели YOLO.....	374
12.3. Использование предварительно обученного детектора RetinaNet.....	378
Итоги главы.....	381
Глава 13. Прогнозирование временных последовательностей.....	382
13.1. Разные виды временных последовательностей.....	382
13.2. Пример прогнозирования температуры.....	383
13.2.1. Подготовка данных.....	387
13.2.2. Базовое решение без привлечения машинного обучения.....	390

13.2.3. Базовое решение с привлечением машинного обучения.....	391
13.2.4. Попытка использовать одномерную сверточную модель.....	393
13.3. Рекуррентные нейронные сети.....	395
13.3.1. Основные особенности рекуррентных нейронных сетей.....	396
13.3.2. Рекуррентный слой Keras.....	399
13.3.3. Приемы эффективного использования рекуррентных нейронных сетей.....	404
13.3.4. Использование рекуррентного прореживания для борьбы с переобучением	404
13.3.5. Наложение нескольких рекуррентных слоев друг на друга	407
13.3.6. Использование двунаправленных рекуррентных нейронных сетей... ..	409
13.4. Что дальше	411
Итоги главы	412
Глава 14. Классификация текста	413
14.1. Краткая история обработки естественных языков.....	413
14.2. Подготовка текстовых данных.....	416
14.2.1. Токенизация на уровнях слов и символов.....	420
14.2.2. Токенизация на уровне субслов	423
14.3. Множества и последовательности.....	428
14.3.1. Загрузка набора данных классификации IMDb.....	429
14.4. Модели множеств.....	431
14.4.1. Обучение модели мешка слов	432
14.4.2. Обучение модели на основе биграмм	436
14.5. Модели последовательностей	437
14.5.1. Обучение рекуррентной модели	439
14.5.2. Векторные представления слов	442
14.5.3. Использование векторных представлений слов.....	444
14.5.4. Предварительное обучение векторного представления слов.....	448
14.5.5. Использование предварительно обученного пространства векторных представлений для классификации	452
Итоги главы	454
Глава 15. Языковые модели и архитектура Transformer	455
15.1. Языковая модель	455
15.1.1. Обучение языковой модели Шекспира	456
15.1.2. Генерирование текста в стиле Шекспира.....	460
15.2. Обучение «последовательность в последовательность»	462
15.2.1. Перевод с английского на испанский	464
15.2.2. Пошаговое обучение с помощью RNN	466
15.3. Архитектура Transformer	471
15.3.1. Скалярное внимание	474
15.3.2. Блок кодировщика модели Transformer	478

15.3.3. Блок декодера в модели Transformer	481
15.3.4. Обучение типа «последовательность в последовательность» модели Transformer.....	483
15.3.5. Внедрение позиционной информации в векторное представление	485
15.4. Классификация с использованием предварительно обученной модели Transformer	489
15.4.1. Предварительно обученный кодировщик модели Transformer.....	489
15.4.2. Загрузка предварительно обученной модели Transformer.....	490
15.4.3. Предварительная обработка отзывов на IMDb.....	493
15.4.4. Дообучение модели Transformer	496
15.5. В чем секрет эффективности архитектуры Transformer?.....	497
Итоги главы.....	500
Глава 16. Генерирование текста.....	502
16.1. Краткая история генерирования последовательностей.....	504
16.2. Обучение миниатюрной модели GPT.....	507
16.2.1. Создание модели.....	510
16.2.2. Предварительное обучение модели.....	512
16.2.3. Генеративное декодирование	515
16.2.4. Стратегии выбора	517
16.3. Использование предварительно обученной LLM.....	521
16.3.1. Генерация текста с использованием модели Gemma	522
16.3.2. Тонкая настройка инструкций.....	525
16.3.3. Низкоранговая адаптация	527
16.4. Дальнейшее развитие больших языковых моделей	532
16.4.1. Обучение с подкреплением на основе отзывов людей.....	532
16.4.2. Мультимодальные модели LLM	536
16.4.3. Генерация с дополненной выборкой.....	539
16.4.4. Рассуждающие модели	541
16.5. В каком направлении продолжат развиваться модели LLM?	543
Итоги главы.....	545
Глава 17. Генерирование изображений.....	547
17.1. Глубокое обучение и генерирование изображений	547
17.1.1. Выбор шаблонов из скрытых пространств изображений.....	548
17.1.2. Вариационные автокодировщики.....	549
12.4.4. Реализация VAE в Keras	552
17.2. Диффузионные модели	558
17.2.1. Набор данных Oxford Flowers.....	560
17.2.2. U-образная сеть шумоподавляющего автокодировщика	561
17.2.3. Понятия времени и графика диффузии	563
17.2.4. Процесс обучения.....	565

17.2.5. Процесс генерации	567
17.2.6. Визуализация результатов с помощью обратного вызова	568
17.2.7. Приступим!	568
17.3. Модели преобразования текста в изображение	570
17.3.1. Исследование скрытого пространства модели преобразования текста в изображение	573
Итоги главы	576
Глава 18. Методы и приемы для применения на практике	578
18.1. Получение максимальной отдачи от моделей	579
18.1.1. Оптимизация гиперпараметров	579
18.1.2. Ансамблирование моделей	586
18.2. Масштабирование обучения моделей	588
18.2.1. Обучение на нескольких GPU	588
18.2.2. Распределенное обучение на практике	591
18.2.3. Обучение на TPU	596
18.3. Ускорение обучения на GPU со смешанной точностью	597
18.3.1. Точность представления чисел с плавающей запятой	597
18.3.2. Прогнозирование с точностью float16	599
18.3.3. Обучение со смешанной точностью	600
18.3.4. Масштабирование потерь в вычислениях со смешанной точностью	600
18.3.5. За гранью смешанной точности: обучение с точностью float8	601
18.3.6. Ускорение прогнозирования с помощью квантования	602
Итоги главы	604
Глава 19. Будущее ИИ	605
19.1. Ограничения глубокого обучения	605
19.1.1. Трудности с адаптацией к новым условиям	605
19.1.2. Чувствительность к формулировкам и несущественным деталям	608
19.1.3. Трудности с обучением обобщаемым программам	610
19.1.4. Риск очеловечивания моделей глубокого обучения	611
19.2. Масштаб — это не все, что нужно	612
19.2.1. Автоматы и носители интеллекта	613
19.2.2. Локальное и экстремальное обобщение	615
19.2.3. Назначение интеллекта	617
19.2.4. Восхождение по спектру обобщения	618
19.3. Реализация интеллекта	619
19.3.1. Гипотеза калейдоскопа	619
19.3.2. Сущность интеллекта: формирование и комбинирование абстракций	620
19.3.3. О важности постановки верной цели	621
19.3.4. Новая цель: адаптация на лету	623

19.3.5. ARC Prize Foundation.....	625
19.3.6. Эра адаптации во время тестирования.....	625
19.3.7. ARC-AGI 2	626
19.4. Недостающие ингредиенты: поиск и символы	628
19.4.1. Два полюса абстракции	628
19.4.2. Познание как соединение обоих видов абстракции	631
19.4.3. Почему глубокое обучение не является исчерпывающим решением задачи генерации абстракций	632
19.4.4. Альтернативный подход к ИИ: синтез программ	633
19.4.5. Сочетание глубокого обучения и синтеза программ.....	634
19.4.6. Непрерывное обучение и комбинирование модульных компонентов.....	637
19.4.7. Долгосрочная перспектива.....	638
Глава 20. Заключение	640
20.1. Краткий обзор ключевых понятий	640
20.1.1. Разные подходы к ИИ.....	640
20.1.2. Что делает глубокое обучение особенным среди других подходов к машинному обучению	641
20.1.3. Как правильно воспринимать глубокое обучение	642
20.1.4. Ключевые технологии.....	643
20.1.5. Обобщенный процесс машинного обучения.....	644
20.1.6. Основные архитектуры сетей.....	645
20.2. Ограничения глубокого обучения.....	650
20.3. Что нас ждет впереди	652
20.4. Как не отстать от прогресса в быстро развивающейся области	653
20.4.1. Практические решения реальных задач на сайте Kaggle.....	653
20.4.2. Знакомство с последними разработками на сайте arXiv	653
20.4.3. Исследование экосистемы Keras	654
Заключительное слово.....	654