

СОДЕРЖАНИЕ

Предисловие	16
Благодарности	17
Об этой книге	19
Об авторах	23
Об иллюстрации на обложке	24
ЧАСТЬ I. ОСНОВЫ ГЛУБОКОГО ОБУЧЕНИЯ	25
Глава 1. Что такое глубокое обучение?	26
1.1. Искусственный интеллект, машинное и глубокое обучение	27
1.1.1. Искусственный интеллект	27
1.1.2. Машинное обучение	28
1.1.3. Обучение представлению данных	29
1.1.4. «Глубина» глубокого обучения	31
1.1.5. Принцип действия глубокого обучения в трех картинках	33
1.1.6. Какой ступени развития достигло глубокое обучение	35
1.1.7. Не верьте рекламе	36
1.1.8. Перспективы ИИ	37
1.2. Что было до глубокого обучения: краткая история машинного обучения	38
1.2.1. Вероятностное моделирование	39
1.2.2. Первые нейронные сети	39

1.2.3. Ядерные методы	40
1.2.4. Деревья решений, случайные леса и градиентный бустинг	41
1.2.5. Назад к нейронным сетям	42
1.2.6. Отличительные черты глубокого обучения.	43
1.2.7. Современный ландшафт машинного обучения	44
1.3. Почему глубокое обучение? Почему сейчас?	45
1.3.1. Оборудование	45
1.3.2. Данные	46
1.3.3. Алгоритмы.	47
1.3.4. Новая волна инвестиций	48
1.3.5. Демократизация глубокого обучения	49
1.3.6. Ждать ли продолжения этой тенденции?	49

Глава 2. Прежде чем начать: математические основы нейронных сетей 51

2.1. Первое знакомство с нейронной сетью	52
2.2. Представление данных для нейронных сетей	56
2.2.1. Скаляры (тензоры нулевого ранга).	57
2.2.2. Векторы (тензоры первого ранга)	57
2.2.3. Матрицы (тензоры второго ранга)	57
2.2.4. Тензоры третьего и высшего рангов.	58
2.2.5. Ключевые атрибуты.	58
2.2.6. Манипулирование тензорами в \mathbb{R}	59
2.2.7. Пакеты данных	60
2.2.8. Практические примеры тензоров с данными.	60
2.2.9. Векторные данные.	61
2.2.10. Временные ряды, или последовательности.	61
2.2.11. Изображения	62
2.2.12. Видео	63

2.3. Шестеренки нейронных сетей: операции с тензорами	63
2.3.1. Поэлементные операции	64
2.3.2. Операции с тензорами, имеющими разные размерности.	65
2.3.3. Скалярное произведение тензоров.	65
2.3.4. Изменение формы тензора.	68
2.3.5. Геометрическая интерпретация операций с тензорами.	69
2.3.6. Геометрическая интерпретация глубокого обучения.	70
2.4. Механизм нейронных сетей: оптимизация на основе градиента	71
2.4.1. Что такое производная?	72
2.4.2. Производная операций с тензорами: градиент	73
2.4.3. Стохастический градиентный спуск	74
2.4.4. Объединение производных: алгоритм обратного распространения ошибки	77
2.5. Взгляд назад на первый пример	78
Краткие итоги главы	80

Глава 3. Начало работы с нейронными сетями 81

3.1. Анатомия нейронной сети	82
3.1.1. Слои: строительные блоки глубокого обучения	83
3.1.2. Модели: сети слоев	84
3.1.3. Функции потерь и оптимизаторы: ключи к настройке процесса обучения	84
3.2. Введение в Keras.	85
3.2.1. Keras, TensorFlow, Theano и CNTK.	86
3.2.2. Установка Keras	87
3.2.3. Разработка с использованием Keras: краткий обзор	88
3.3. Настройка рабочей станции для глубокого обучения	90
3.3.1. Подготовка Keras: два варианта	90
3.3.2. Запуск заданий глубокого обучения в облаке: за и против	91
3.3.3. Выбор GPU для глубокого обучения	91

3.4. Классификация отзывов к фильмам: пример бинарной классификации	92
3.4.1. Набор данных IMDb	92
3.4.2. Подготовка данных	94
3.4.3. Конструирование сети	94
3.4.4. Проверка решения	98
3.4.5. Использование обученной сети для предсказаний на новых данных	102
3.4.6. Дальнейшие эксперименты	102
3.4.7. Итоги	103
3.5. Классификация новостных лент: пример классификации в несколько классов	103
3.5.1. Набор данных Reuters	104
3.5.2. Подготовка данных	105
3.5.3. Конструирование сети	106
3.5.4. Проверка решения	107
3.5.5. Предсказания на новых данных	108
3.5.6. Другой способ обработки меток и потерь	109
3.5.7. Важность использования достаточно больших промежуточных уровней	109
3.5.8. Дальнейшие эксперименты	110
3.5.9. Итоги	110
3.6. Предсказание цен на дома: пример регрессии	111
3.6.1. Набор данных с ценами на жилье в Бостоне	111
3.6.2. Подготовка данных	112
3.6.3. Конструирование сети	113
3.6.4. Оценка решения методом перекрестной проверки по K блокам	114
3.6.5. Подведение итогов	118
Краткие итоги главы	118

Глава 4. Основы машинного обучения	120
4.1. Четыре раздела машинного обучения	120
4.1.1. Контролируемое обучение	121
4.1.2. Неконтролируемое обучение	121
4.1.3. Самоконтролируемое обучение	122
4.1.4. Обучение с подкреплением	122
4.2. Оценка моделей машинного обучения	124
4.2.1. Тренировочные, проверочные и контрольные наборы данных	124
4.2.2. Что важно помнить	128
4.3. Обработка данных, конструирование признаков и обучение признаков	128
4.3.1. Предварительная обработка данных для нейронных сетей	129
4.3.2. Конструирование признаков	131
4.4. Переобучение и недообучение	133
4.4.1. Уменьшение размера сети	133
4.4.2. Добавление регуляризации весов	137
4.4.3. Добавление прореживания	138
4.5. Обобщенный процесс решения задач машинного обучения	141
4.5.1. Определение задачи и создание набора данных	141
4.5.2. Выбор меры успеха	142
4.5.3. Выбор протокола оценки	143
4.5.4. Предварительная подготовка данных	143
4.5.5. Разработка модели более совершенной, чем базовый случай	144
4.5.6. Масштабирование по вертикали: разработка модели с переобучением	145
4.5.7. Регуляризация модели и настройка гиперпараметров	146
Краткие итоги главы	147

ЧАСТЬ II. ГЛУБОКОЕ ОБУЧЕНИЕ НА ПРАКТИКЕ 149**Глава 5. Глубокое обучение для распознавания образов. 150**

5.1. Введение в сверточные нейронные сети	150
5.1.1. Операция свертывания	153
5.1.2. Выбор максимального значения из соседних (max-pooling)	159
5.2. Обучение сверточной нейронной сети с нуля на небольшом наборе данных	161
5.2.1. Целесообразность глубокого обучения для решения задач с небольшими наборами данных	162
5.2.2. Загрузка данных	162
5.2.3. Конструирование сети	165
5.2.4. Предварительная обработка данных	167
5.2.5. Расширение данных	170
5.3. Использование предварительно обученной сверточной нейронной сети.	174
5.3.1. Выделение признаков	175
5.3.2. Дообучение	185
5.3.3. Итоги.	190
5.4. Визуализация знаний, заключенных в сверточной нейронной сети.	190
5.4.1. Визуализация промежуточных активаций.	191
5.4.2. Визуализация фильтров сверточных нейронных сетей	198
5.4.3. Визуализация тепловых карт активации класса	203
Краткие итоги главы	208

**Глава 6. Глубокое обучение для текста
и последовательностей 209**

6.1. Работа с текстовыми данными.	210
6.1.1. Прямое кодирование слов и символов	212
6.1.2. Использование векторного представления слов	214

6.1.3. Объединение всего вместе: от исходного текста к векторному представлению слов	221
6.1.4. Итоги.	227
6.2. Рекуррентные нейронные сети	228
6.2.1. Рекуррентный уровень в Keras	231
6.2.2. Слои LSTM и GRU.	234
6.2.3. Пример использования слоя LSTM из Keras.	238
6.2.4. Подведение итогов	239
6.3. Улучшенные методы использования рекуррентных нейронных сетей.	240
6.3.1. Задача прогнозирования температуры	240
6.3.2. Подготовка данных	243
6.3.3. Базовое решение без привлечения машинного обучения	247
6.3.4. Базовое решение с привлечением машинного обучения	248
6.3.5. Первое базовое рекуррентное решение	250
6.3.6. Использование рекуррентного прореживания для борьбы с переобучением	252
6.3.7. Наложение нескольких рекуррентных уровней друг на друга.	254
6.3.8. Использование двунаправленных рекуррентных нейронных сетей	255
6.3.9. Что дальше	260
6.3.10. Подведение итогов	260
6.4. Обработка последовательностей с помощью сверточных нейронных сетей.	261
6.4.1. Обработка последовательных данных с помощью одномерной сверточной нейронной сети	262
6.4.2. Выбор соседних значений в одномерной последовательности данных	263
6.4.3. Реализация одномерной сверточной сети.	263
6.4.4. Объединение сверточных и рекуррентных сетей для обработки длинных последовательностей	265
6.4.5. Подведение итогов	270
Краткие итоги главы	270

Глава 7. Лучшие практики глубокого обучения продвинутого уровня	272
7.1. За рамками последовательной модели: функциональный API фреймворка Keras	273
7.1.1. Введение в функциональный API	276
7.1.2. Модели с несколькими входами	278
7.1.3. Модели с несколькими выходами	280
7.1.4. Ориентированные ациклические графы уровней	283
7.1.5. Повторное использование экземпляров слоев	288
7.1.6. Модели как слои	289
7.1.7. Подведение итогов.	290
7.2. Исследование имониторинг моделей глубокого обучения с использованием обратных вызовов Keras иTensorBoard	290
7.2.1. Применение обратных вызовов для воздействия на модель в ходе обучения	291
7.2.2. Введение в TensorBoard: фреймворк визуализации TensorFlow	294
7.2.3. Подведение итогов	300
7.3. Извлечение максимальной пользы из моделей.	300
7.3.1. Шаблоны улучшенных архитектур	300
7.3.2. Оптимизация гиперпараметров	304
7.3.3. Ансамблирование моделей.	306
7.3.4. Подведение итогов	308
Краткие итоги главы	309
Глава 8. Генеративное глубокое обучение	310
8.1. Генерирование текста с помощью LSTM	312
8.1.1. Краткая история генеративных рекуррентных сетей.	312
8.1.2. Как генерируются последовательности данных?.	313

8.1.3. Важность стратегии выбора	314
8.1.4. Реализация посимвольной генерации текста на основе LSTM	316
8.1.5. Подведение итогов	321
8.2. DeepDream	322
8.2.1. Реализация DeepDream в Keras	323
8.2.2. Подведение итогов	328
8.3. Нейронная передача стиля	329
8.3.1. Функция потерь содержимого	331
8.3.2. Функция потерь стиля	331
8.3.3. Нейронная передача стиля в Keras	332
8.3.4. Подведение итогов	338
8.4. Генерирование изображений с вариационными автокодировщиками	338
8.4.1. Выбор шаблонов из скрытых пространств изображений	340
8.4.2. Концептуальные векторы для редактирования изображений	341
8.4.3. Вариационные автокодировщики	342
8.4.4. Подведение итогов	349
8.5. Введение в генеративно-сопоставительные сети	350
8.5.1. Реализация простейшей генеративно-сопоставительной сети	352
8.5.2. Набор хитростей	352
8.5.3. Генератор	354
8.5.4. Дискриминатор	355
8.5.5. Сопоставительная сеть	355
8.5.6. Как обучить сеть DCGAN	356
8.5.7. Подведение итогов	359
Краткие итоги главы	359

Глава 9. Заключение	360
9.1. Краткий обзор ключевых понятий	361
9.1.1. Разные подходы к ИИ.	361
9.1.2. Что делает глубокое обучение особенным среди других подходов к машинному обучению	361
9.1.3. Как правильно воспринимать глубокое обучение	362
9.1.4. Ключевые технологии	363
9.1.5. Обобщенный процесс машинного обучения	364
9.1.6. Основные архитектуры сетей	365
9.1.7. Пространство возможностей.	370
9.2. Ограничения глубокого обучения	372
9.2.1. Риск очеловечивания моделей глубокого обучения	373
9.2.2. Локальное и экстремальное обобщение	375
9.2.3. Итоги.	377
9.3. Будущее глубокого обучения.	377
9.3.1. Модели как программы	378
9.3.2. За границами алгоритма обратного распространения ошибки и дифференцируемых уровней	380
9.3.3. Автоматизированное машинное обучение	381
9.3.4. Непрерывное обучение и повторное использование модульных подпрограмм	382
9.3.5. Долгосрочная перспектива	384
9.4. Как не отстать от прогресса в быстро развивающейся области	385
9.4.1. Практические решения реальных задач на сайте Kaggle.	385
9.4.2. Знакомство с последними разработками на сайте arXiv.	385
9.4.3. Исследование экосистемы Keras	386
9.5. Заключительное слово	387

Приложение А. Установка Keras и его зависимостей в Ubuntu	388
А.1. Обзор процесса установки	388
А.2. Установка системных библиотек	388
А.3. Настройка поддержки GPU	389
А.3.1. Установка CUDA	389
А.3.2. Установка cuDNN	390
А.3.3. Настройка окружения для CUDA	390
А.4. Установка Keras и TensorFlow	391
Приложение В. Запуск RStudio Server на экземпляре EC2 GPU	393
В.1. Зачем использовать AWS для глубокого обучения?	393
В.2. Когда нежелательно использовать AWS для глубокого обучения?	393
В.3. Настройка экземпляра AWS GPU	394
В.3.1. Установка R и RStudio Server	396
В.3.2. Настройка CUDA	397
В.3.3. Подготовка библиотек для Keras	397
В.4. Настройка доступа к RStudio Server	398
В.5. Установка Keras	399