

Научная библиотека

БНТУ



# Глубокое обучение на Python

Второе международное издание

Франсуа Шолле



 ПИТЕР®

Санкт-Петербург · Москва · Минск

2023

# *Краткое содержание*

---

Предисловие . . . . .	18
Благодарности . . . . .	20
О книге. . . . .	21
Об авторе . . . . .	23
Иллюстрация на обложке . . . . .	24
От издательства . . . . .	25
<b>Глава 1.</b> Что такое глубокое обучение . . . . .	26
<b>Глава 2.</b> Математические основы нейронных сетей. . . . .	56
<b>Глава 3.</b> Введение в Keras и TensorFlow . . . . .	103
<b>Глава 4.</b> Начало работы с нейронными сетями: классификация и регрессия . . . . .	135
<b>Глава 5.</b> Основы машинного обучения. . . . .	165
<b>Глава 6.</b> Обобщенный процесс машинного обучения. . . . .	203
<b>Глава 7.</b> Работа с Keras: глубокое погружение . . . . .	226
<b>Глава 8.</b> Введение в глубокое обучение в технологиях компьютерного зрения . . . . .	260
<b>Глава 9.</b> Продвинутое приемы глубокого обучения в технологиях компьютерного зрения . . . . .	303
<b>Глава 10.</b> Глубокое обучение на временных последовательностях . . . . .	350
<b>Глава 11.</b> Глубокое обучение для текста . . . . .	384
<b>Глава 12.</b> Генеративное глубокое обучение . . . . .	449
<b>Глава 13.</b> Методы и приемы для применения на практике . . . . .	506
<b>Глава 14.</b> Заключение . . . . .	529

# Оглавление

---

Предисловие . . . . .	18
Благодарности . . . . .	20
О книге. . . . .	21
Кому адресована эта книга . . . . .	21
О примерах кода. . . . .	22
Об авторе . . . . .	23
Иллюстрация на обложке . . . . .	24
От издательства . . . . .	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. Изучение правил и представлений данных. . . . .	30
1.1.4. «Глубина» глубокого обучения . . . . .	33
1.1.5. Принцип действия глубокого обучения в трех картинках . . . . .	35
1.1.6. Какой ступени развития достигло глубокое обучение. . . . .	37
1.1.7. Не верьте рекламе . . . . .	38
1.1.8. Перспективы ИИ . . . . .	39
1.2. Что было до глубокого обучения: краткая история машинного обучения . . . . .	40
1.2.1. Вероятностное моделирование . . . . .	40
1.2.2. Первые нейронные сети. . . . .	41

1.2.3. Ядерные методы . . . . .	41
1.2.4. Деревья решений, случайные леса и градиентный бустинг . . . . .	43
1.2.5. Назад к нейронным сетям . . . . .	44
1.2.6. Отличительные черты глубокого обучения . . . . .	45
1.2.7. Современный ландшафт машинного обучения . . . . .	46
1.3. Почему глубокое обучение? Почему сейчас? . . . . .	49
1.3.1. Оборудование . . . . .	49
1.3.2. Данные . . . . .	51
1.3.3. Алгоритмы . . . . .	51
1.3.4. Новая волна инвестиций . . . . .	52
1.3.5. Демократизация глубокого обучения . . . . .	54
1.3.6. Ждать ли продолжения этой тенденции? . . . . .	54
<b>Глава 2. Математические основы нейронных сетей. . . . .</b>	<b>56</b>
2.1. Первое знакомство с нейронной сетью . . . . .	57
2.2. Представление данных для нейронных сетей . . . . .	61
2.2.1. Скаляры (тензоры нулевого ранга) . . . . .	61
2.2.2. Векторы (тензоры первого ранга) . . . . .	62
2.2.3. Матрицы (тензоры второго ранга) . . . . .	62
2.2.4. Тензоры третьего и более высоких рангов . . . . .	62
2.2.5. Ключевые атрибуты . . . . .	63
2.2.6. Манипулирование тензорами с помощью NumPy . . . . .	64
2.2.7. Пакеты данных . . . . .	65
2.2.8. Практические примеры тензоров с данными. . . . .	66
2.2.9. Векторные данные . . . . .	66
2.2.10. Временные ряды или последовательности . . . . .	67
2.2.11. Изображения. . . . .	67
2.2.12. Видео. . . . .	68
2.3. Шестеренки нейронных сетей: операции с тензорами. . . . .	69
2.3.1. Поэлементные операции . . . . .	70
2.3.2. Расширение . . . . .	71
2.3.3. Скалярное произведение тензоров . . . . .	73
2.3.4. Изменение формы тензора . . . . .	75
2.3.5. Геометрическая интерпретация операций с тензорами . . . . .	76
2.3.6. Геометрическая интерпретация глубокого обучения . . . . .	80

2.4. Механизм нейронных сетей: оптимизация на основе градиента . . . . .	81
2.4.1. Что такое производная . . . . .	83
2.4.2. Производная операций с тензорами: градиент. . . . .	84
2.4.3. Стохастический градиентный спуск . . . . .	86
2.4.4. Объединение производных: алгоритм обратного распространения ошибки . . . . .	89
2.5. Оглядываясь на первый пример . . . . .	95
2.5.1. Повторная реализация первого примера в TensorFlow . . . . .	97
2.5.2. Выполнение одного этапа обучения . . . . .	99
2.5.3. Полный цикл обучения . . . . .	100
2.5.4. Оценка модели . . . . .	101
Краткие итоги главы . . . . .	101
<b>Глава 3. Введение в Keras и TensorFlow . . . . .</b>	<b>103</b>
3.1. Что такое TensorFlow . . . . .	104
3.2. Что такое Keras . . . . .	105
3.3. Keras и TensorFlow: краткая история . . . . .	106
3.4. Настройка окружения для глубокого обучения. . . . .	107
3.4.1. Jupyter Notebook: предпочтительный способ проведения экспериментов с глубоким обучением . . . . .	109
3.4.2. Использование Colaboratory. . . . .	109
3.5. Первые шаги с TensorFlow . . . . .	112
3.5.1. Тензоры-константы и тензоры-переменные . . . . .	113
3.5.2. Операции с тензорами: математические действия в TensorFlow . . . . .	115
3.5.3. Второй взгляд на GradientTape . . . . .	116
3.5.4. Полный пример: линейный классификатор на TensorFlow. . . . .	117
3.6. Анатомия нейронной сети: знакомство с основами Keras. . . . .	122
3.6.1. Слои: строительные блоки глубокого обучения. . . . .	122
3.6.2. От слоев к моделям . . . . .	126
3.6.3. Этап «компиляции»: настройка процесса обучения . . . . .	128
3.6.4. Выбор функции потерь . . . . .	130
3.6.5. Метод fit() . . . . .	130
3.6.6. Оценка потерь и метрик на проверочных данных. . . . .	131
3.6.7. Вывод: использование модели после обучения . . . . .	133
Краткие итоги главы . . . . .	134

<b>Глава 4. Начало работы с нейронными сетями: классификация и регрессия</b> . . . . .	135
4.1. Классификация отзывов к фильмам: пример бинарной классификации. . . . .	137
4.1.1. Набор данных IMDb . . . . .	137
4.1.2. Подготовка данных. . . . .	139
4.1.3. Конструирование модели. . . . .	140
4.1.4. Проверка решения . . . . .	143
4.1.5. Использование обученной сети для предсказаний на новых данных. . . . .	146
4.1.6. Дальнейшие эксперименты . . . . .	147
4.1.7. Подведение итогов . . . . .	147
4.2. Классификация новостных лент: пример классификации в несколько классов. . . . .	148
4.2.1. Набор данных Reuters . . . . .	148
4.2.2. Подготовка данных. . . . .	149
4.2.3. Конструирование модели. . . . .	150
4.2.4. Проверка решения . . . . .	151
4.2.5. Предсказания на новых данных. . . . .	153
4.2.6. Другой способ обработки меток и потерь. . . . .	154
4.2.7. Важность использования достаточно больших промежуточных слоев . . . . .	154
4.2.8. Дальнейшие эксперименты . . . . .	155
4.2.9. Подведение итогов . . . . .	155
4.3. Предсказание цен на дома: пример регрессии . . . . .	156
4.3.1. Набор данных с ценами на жилье в Бостоне . . . . .	156
4.3.2. Подготовка данных. . . . .	157
4.3.3. Конструирование модели. . . . .	158
4.3.4. Оценка решения методом перекрестной проверки по K блокам . . . . .	159
4.3.5. Предсказания на новых данных. . . . .	163
4.3.6. Подведение итогов . . . . .	163
Краткие итоги главы . . . . .	163
<b>Глава 5. Основы машинного обучения</b> . . . . .	165
5.1. Обобщение: цель машинного обучения . . . . .	165
5.1.1. Недообучение и переобучение. . . . .	166
5.1.2. Природа общности в глубоком обучении. . . . .	172

5.2. Оценка моделей машинного обучения . . . . .	180
5.2.1. Обучающие, проверочные и контрольные наборы данных . . . . .	180
5.2.2. Выбор базового уровня . . . . .	184
5.2.3. Что важно помнить об оценке моделей . . . . .	185
5.3. Улучшение качества обучения модели . . . . .	185
5.3.1. Настройка основных параметров градиентного спуска . . . . .	186
5.3.2. Использование более удачной архитектуры . . . . .	187
5.3.3. Увеличение емкости модели . . . . .	188
5.4. Улучшение общности . . . . .	190
5.4.1. Курирование набора данных . . . . .	190
5.4.2. Конструирование признаков . . . . .	191
5.4.3. Ранняя остановка . . . . .	193
5.4.4. Регуляризация модели . . . . .	193
Краткие итоги главы . . . . .	202
<b>Глава 6. Обобщенный процесс машинного обучения . . . . .</b>	<b>203</b>
6.1. Определение задачи . . . . .	205
6.1.1. Формулировка задачи . . . . .	205
6.1.2. Сбор данных . . . . .	207
6.1.3. Первичный анализ данных . . . . .	211
6.1.4. Выбор меры успеха . . . . .	212
6.2. Разработка модели . . . . .	212
6.2.1. Подготовка данных . . . . .	213
6.2.2. Выбор протокола оценки . . . . .	215
6.2.3. Преодоление базового случая . . . . .	215
6.2.4. Следующий шаг: разработка модели с переобучением . . . . .	217
6.2.5. Регуляризация и настройка модели . . . . .	218
6.3. Развертывание модели . . . . .	219
6.3.1. Объяснение особенностей работы модели заинтересованным сторонам и обозначение границ ожидаемого . . . . .	219
6.3.2. Предоставление доступа к модели . . . . .	220
6.3.3. Мониторинг качества работы модели в процессе эксплуатации . . . . .	223
6.3.4. Обслуживание модели . . . . .	224
Краткие итоги главы . . . . .	225

<b>Глава 7. Работа с Keras: глубокое погружение</b> . . . . .	226
7.1. Спектр рабочих процессов . . . . .	227
7.2. Разные способы создания моделей Keras . . . . .	227
7.2.1. Последовательная модель Sequential . . . . .	228
7.2.2. Функциональный API . . . . .	231
7.2.3. Создание производных от класса Model . . . . .	239
7.2.4. Смешивание и согласование различных компонентов. . . . .	241
7.2.5. Используйте правильный инструмент . . . . .	242
7.3. Встроенные циклы обучения и оценки . . . . .	243
7.3.1. Использование собственных метрик. . . . .	244
7.3.2. Использование обратных вызовов . . . . .	245
7.3.3. Разработка своего обратного вызова . . . . .	247
7.3.4. Мониторинг и визуализация с помощью TensorBoard . . . . .	249
7.4. Разработка своего цикла обучения и оценки. . . . .	251
7.4.1. Обучение и прогнозирование . . . . .	252
7.4.2. Низкоуровневое использование метрик . . . . .	253
7.4.3. Полный цикл обучения и оценки . . . . .	254
7.4.4. Ускорение вычислений с помощью tf.function. . . . .	256
7.4.5. Использование fit() с нестандартным циклом обучения. . . . .	257
Краткие итоги главы . . . . .	259
<b>Глава 8. Введение в глубокое обучение в технологиях компьютерного зрения</b> . . . . .	260
8.1. Введение в сверточные нейронные сети. . . . .	261
8.1.1. Операция свертывания . . . . .	264
8.1.2. Выбор максимального значения из соседних (max-pooling) . . . . .	269
8.2. Обучение сверточной нейронной сети с нуля на небольшом наборе данных . . . . .	272
8.2.1. Целесообразность глубокого обучения для решения задач с небольшими наборами данных . . . . .	272
8.2.2. Загрузка данных . . . . .	273
8.2.3. Конструирование сети. . . . .	276
8.2.4. Предварительная обработка данных . . . . .	278
8.2.5. Обогащение данных . . . . .	283



8.3. Использование предварительно обученной модели . . . . .	288
8.3.1. Выделение признаков . . . . .	289
8.3.2. Дообучение предварительно обученной модели . . . . .	298
Краткие итоги главы . . . . .	302
<b>Глава 9. Продвинутое обучение в технологиях компьютерного зрения . . . . .</b>	<b>303</b>
9.1. Три основные задачи в сфере компьютерного зрения . . . . .	303
9.2. Пример сегментации изображения. . . . .	305
9.3. Современные архитектурные шаблоны сверточных сетей . . . . .	313
9.3.1. Модульность, иерархия, многократное использование . . . . .	314
9.3.2. Остаточные связи . . . . .	317
9.3.3. Пакетная нормализация . . . . .	321
9.3.4. Раздельная свертка по глубине . . . . .	324
9.3.5. Собираем все вместе: мини-модель с архитектурой Xception . . . . .	326
9.4. Интерпретация знаний, заключенных в сверточной нейронной сети. . . . .	329
9.4.1. Визуализация промежуточных активаций . . . . .	330
9.4.2. Визуализация фильтров сверточных нейронных сетей . . . . .	337
9.4.3. Визуализация тепловых карт активации класса. . . . .	343
Краткие итоги главы . . . . .	349
<b>Глава 10. Глубокое обучение на временных последовательностях . . . . .</b>	<b>350</b>
10.1. Разные виды временных последовательностей . . . . .	350
10.2. Пример прогнозирования температуры . . . . .	352
10.2.1. Подготовка данных . . . . .	355
10.2.2. Базовое решение без привлечения машинного обучения . . . . .	359
10.2.3. Базовое решение с привлечением машинного обучения . . . . .	360
10.2.4. Попытка использовать одномерную сверточную модель . . . . .	362
10.2.5. Первое базовое рекуррентное решение . . . . .	364
10.3. Рекуррентные нейронные сети . . . . .	366
10.3.1. Рекуррентный слой в Keras. . . . .	369

10.4. Улучшенные методы использования рекуррентных нейронных сетей . . . . .	373
10.4.1. Использование рекуррентного прореживания для борьбы с переобучением . . . . .	374
10.4.2. Наложение нескольких рекуррентных слоев друг на друга . . . . .	377
10.4.3. Использование двунаправленных рекуррентных нейронных сетей . . . . .	379
10.4.4. Что дальше . . . . .	382
Краткие итоги главы . . . . .	383
<b>Глава 11. Глубокое обучение для текста . . . . .</b>	<b>384</b>
11.1. Обработка естественных языков . . . . .	384
11.2. Подготовка текстовых данных . . . . .	387
11.2.1. Стандартизация текста . . . . .	388
11.2.2. Деление текста на единицы (токенизация) . . . . .	389
11.2.3. Индексирование словаря . . . . .	390
11.2.4. Использование слоя TextVectorization . . . . .	392
11.3. Два подхода к представлению групп слов: множества и последовательности . . . . .	396
11.3.1. Подготовка данных IMDb с отзывами к фильмам . . . . .	397
11.3.2. Обработка наборов данных: мешки слов . . . . .	399
11.3.3. Обработка слов как последовательностей: модели последовательностей . . . . .	406
11.4. Архитектура Transformer . . . . .	417
11.4.1. Идея внутреннего внимания . . . . .	417
11.4.2. Многоголовое внимание . . . . .	423
11.4.3. Кодировщик Transformer . . . . .	424
11.4.4. Когда использовать модели последовательностей вместо моделей мешка слов . . . . .	431
11.5. За границами классификации текста: обучение «последовательность в последовательность» . . . . .	432
11.5.1. Пример машинного перевода . . . . .	434
11.5.2. Обучение типа «последовательность в последовательность» рекуррентной сети . . . . .	437
11.5.3. Обучение типа «последовательность в последовательность» архитектуры Transformer . . . . .	442
Краткие итоги главы . . . . .	448

Глава 12. Генеративное глубокое обучение . . . . .	449
12.1. Генерирование текста . . . . .	451
12.1.1. Краткая история генеративного глубокого обучения для генерирования последовательностей . . . . .	451
12.1.2. Как генерируются последовательности данных . . . . .	452
12.1.3. Важность стратегии выбора . . . . .	453
12.1.4. Реализация генерации текста в Keras . . . . .	456
12.1.5. Обратный вызов для генерации текста с разными значениями температуры . . . . .	460
12.1.6. Подведение итогов . . . . .	463
12.2. DeepDream . . . . .	464
12.2.1. Реализация DeepDream в Keras . . . . .	465
12.2.2. Подведение итогов . . . . .	472
12.3. Нейронная передача стиля . . . . .	473
12.3.1. Функция потерь содержимого . . . . .	474
12.3.2. Функция потерь стиля . . . . .	474
12.3.3. Нейронная передача стиля в Keras . . . . .	475
12.3.4. Подведение итогов . . . . .	481
12.4. Генерирование изображений с вариационными автокодировщиками . . . . .	482
12.4.1. Выбор шаблонов из скрытых пространств изображений . . . . .	482
12.4.2. Концептуальные векторы для редактирования изображений . . . . .	483
12.4.3. Вариационные автокодировщики . . . . .	484
12.4.4. Реализация VAE в Keras . . . . .	487
12.4.5. Подведение итогов . . . . .	493
12.5. Введение в генеративно-состязательные сети . . . . .	493
12.5.1. Реализация простейшей генеративно-состязательной сети . . . . .	495
12.5.2. Набор хитростей . . . . .	496
12.5.3. Получение набора данных CelebA . . . . .	497
12.5.4. Дискриминатор . . . . .	498
12.5.5. Генератор . . . . .	499
12.5.6. Состязательная сеть . . . . .	501
12.5.7. Подведение итогов . . . . .	504
Краткие итоги главы . . . . .	505

<b>Глава 13. Методы и приемы для применения на практике</b> . . . . .	506
13.1. Получение максимальной отдачи от моделей . . . . .	507
13.1.1. Оптимизация гиперпараметров . . . . .	507
13.1.2. Ансамблирование моделей . . . . .	515
13.2. Масштабирование обучения моделей . . . . .	517
13.2.1. Ускорение обучения на GPU со смешанной точностью. . . . .	518
13.2.2. Обучение на нескольких GPU . . . . .	522
13.2.3. Обучение на TPU . . . . .	525
Краткие итоги главы . . . . .	528
<b>Глава 14. Заключение</b> . . . . .	529
14.1. Краткий обзор ключевых понятий . . . . .	530
14.1.1. Разные подходы к ИИ . . . . .	530
14.1.2. Что делает глубокое обучение особенным среди других подходов к машинному обучению . . . . .	531
14.1.3. Как правильно воспринимать глубокое обучение . . . . .	531
14.1.4. Ключевые технологии . . . . .	533
14.1.5. Обобщенный процесс машинного обучения . . . . .	534
14.1.6. Основные архитектуры сетей . . . . .	535
14.1.7. Пространство возможностей . . . . .	540
14.2. Ограничения глубокого обучения . . . . .	542
14.2.1. Риск очеловечивания моделей глубокого обучения . . . . .	543
14.2.2. Автоматы и носители интеллекта . . . . .	546
14.2.3. Локальное и экстремальное обобщение . . . . .	548
14.2.4. Назначение интеллекта . . . . .	550
14.2.5. Восхождение по спектру обобщения . . . . .	551
14.3. Курс на увеличение универсальности в ИИ . . . . .	552
14.3.1. О важности постановки верной цели: правило выбора кратчайшего пути . . . . .	553
14.3.2. Новая цель . . . . .	555
14.4. Реализация интеллекта: недостающие ингредиенты . . . . .	557
14.4.1. Интеллект как чувствительность к абстрактным аналогиям . . . . .	557
14.4.2. Два полюса абстракции . . . . .	559
14.4.3. Недостающая половина картины . . . . .	563

14.5. Будущее глубокого обучения . . . . .	564
14.5.1. Модели как программы . . . . .	565
14.5.2. Сочетание глубокого обучения и синтеза программ . . . . .	566
14.5.3. Непрерывное обучение и повторное использование модульных подпрограмм. . . . .	569
14.5.4. Долгосрочная перспектива . . . . .	571
14.6. Как не отстать от прогресса в быстроразвивающейся области . . . . .	572
14.6.1. Практические решения реальных задач на сайте Kaggle. . . . .	572
14.6.2. Знакомство с последними разработками на сайте arXiv . . . . .	573
14.6.3. Исследование экосистемы Keras . . . . .	573
Заключительное слово . . . . .	574