

Оглавление

Предисловие от издательства	15
Предисловие	16
Обзор книги	18
Часть I	18
Часть II	19
Часть III	20
Часть IV	21
Часть V	21
Библиографические примечания	22
Благодарности	24
Диаграмма глав	26
Символы и обозначения	27
ЧАСТЬ I. ВВЕДЕНИЕ В НЕЙРОННЫЕ СЕТИ	31
Глава 1. Вводные задачи	32
1.1. Вода в резервуаре	32
1.2. Электронная схема	34
1.3. Задача о восьми ладьях	36
1.4. Биологический нейрон	37
1.5. Линейная регрессия	39
1.6. Сеть – фабрика по производству коктейлей	42
1.7. Электронная сеть	43
1.8. В заключение	45
1.9. Упражнения	46
Глава 2. Функции активации	48
2.1. Примеры функций активации	48
2.2. Сигмоидные функции	58
2.3. Сжимающие функции	62
2.4. В заключение	63
2.5. Упражнения	64
Глава 3. Функции стоимости	66
3.1. Входные данные, выходные данные и цель	66
3.2. Функция супремума ошибки	67
3.3. Функция ошибки L^2	67

3.4. Функция среднеквадратической ошибки	68
3.5. Перекрестная энтропия.....	71
3.6. Расстояние Кульбака–Лейблера	72
3.7. Расстояние Дженсена–Шеннона	75
3.8. Максимальное среднее расхождение	76
3.9. Другие функции стоимости.....	79
3.10. Выборочная оценка функций стоимости	79
3.11. Функции стоимости и регуляризация	82
3.12. Ошибки обучения и тестирования	83
3.13. Геометрический смысл	86
3.14. В заключение	89
3.15. Упражнения	89

Глава 4. Алгоритмы поиска минимума 92

4.1. Общие свойства минимумов.....	92
4.1.1. Функции действительной переменной.....	92
4.1.2. Функции нескольких действительных переменных	93
4.2. Алгоритм градиентного спуска.....	96
4.2.1. Наборы уровней	96
4.2.3. Метод наискорейшего спуска	103
4.3. Кинематическая интерпретация	110
4.4. Метод импульса	113
4.4.1. Кинематическая интерпретация	114
4.4.2. Условия сходимости	117
4.5. AdaGrad	119
4.6. RMSProp	120
4.7. Adam	122
4.8. AdaMax	123
4.9. Метод имитации отжига.....	124
4.9.1. Кинематическая интерпретация метода имитации отжига.....	124
4.9.2. Термодинамическая интерпретация метода имитации отжига ...	126
4.10. Метод увеличения разрешения.....	128
4.11. Метод Гессе	131
4.12. Метод Ньютона.....	133
4.13. Стохастический поиск	134
4.13.1. Детерминированный вариант.....	134
4.13.2. Стохастический вариант.....	135
4.14. Поиск в окрестности	138
4.14.1. Поиск слева и справа	139
4.14.2. Поиск по окружности.....	139
4.14.3. Стохастический поиск в сфере.....	141
4.14.4. От локального к глобальному.....	142
4.15. Непрерывное обучение	143
4.16. В заключение	145
4.17. Упражнения	146

Глава 5. Абстрактные нейроны 148

5.1. Определение и свойства.....	148
----------------------------------	-----

5.2. Модель перцептрона.....	150
5.3. Сигмоидный нейрон.....	156
5.4. Логистическая регрессия.....	158
5.4.1. Вероятность дефолта компании	158
5.4.2. Бинарный классификатор	161
5.4.3. Обучение с использованием квадратической разности как функции стоимости	165
5.5. Линейный нейрон	166
5.6. Адалин	170
5.7. Мадалин	171
5.8. Нейрон с континуальным входом	172
5.9. В заключение	175
5.10. Упражнения	175
Глава 6. Нейронные сети.....	178
6.1. Пример нейронной сети.....	178
6.1.1. Полная вариация и регуляризация	182
6.1.2. Обратное распространение ошибки.....	183
6.2. Общие нейронные сети	187
6.2.1. Прямое распространение через сеть	189
6.2.2. Обратное распространение через сеть.....	189
6.2.3. Обратное распространение дельт	190
6.2.4. В заключение об отношениях	192
6.2.5. Матричная форма	192
6.2.6. Алгоритм градиентного спуска.....	195
6.2.7. Проблема исчезающего градиента	195
6.2.8. Пакетное обучение.....	197
6.2.9. Определение нейронной сети прямого распространения.....	198
6.3. Инициализация весов.....	199
6.4. Сильные и слабые априоры.....	203
6.5. В заключение	204
6.6. Упражнения	205
ЧАСТЬ II. АНАЛИТИЧЕСКАЯ ТЕОРИЯ	207
Глава 7. Аппроксимационные теоремы	208
7.1. Теорема Дини	208
7.2. Теорема Арцела–Асколи	210
7.3. Применение к нейронным сетям.....	213
7.4. Теорема Стоуна–Вейерштрасса.....	214
7.5. Применение к нейронным сетям.....	216
7.6. Тауберовы теоремы Винера.....	218
7.6.1. Изучение сигналов в $L^1(\mathbb{R})$	218
7.6.2. Изучение сигналов в $L^2(\mathbb{R})$	219
7.7. Принцип сжатия	220
7.8. Приложение к рекуррентным нейронным сетям	222
7.9. Резонансные сети	226

7.10. В заключение	227
7.11. Упражнения	228

Глава 8. Обучение на одномерных входных данных..... 230

8.1. Предварительные сведения	230
8.2. Перцептронная сеть с одним скрытым слоем	232
8.3. Сигмоидная сеть с одним скрытым слоем.....	234
8.4. Обучение с использованием функции ReLU	239
8.5. Обучение с использованием функции softplus.....	243
8.6. Дискретные данные	248
8.7. В заключение	249
8.8. Упражнения	250

Глава 9. Универсальные аппроксиматоры..... 252

9.1. Вводные примеры.....	252
9.2. Общие положения.....	253
9.3. Сети с одним скрытым слоем	255
9.3.1. Изучение непрерывных функций $f \in C(I_n)$	255
9.3.2. Изучение квадратично интегрируемых функций $f \in L^2(I_n)$	260
9.3.3. Изучение интегрируемых функций $f \in L^1(I_n)$	265
9.3.4. Изучение измеримых функций $f \in \mathcal{M}(\mathbb{R}^n)$	267
9.4. Оценка границ ошибки	275
9.5. Изучение q-интегрируемых функций, $f \in L^q(\mathbb{R}^n)$	276
9.6. Изучение решений ОДУ.....	278
9.7. В заключение	280
9.8. Упражнения	280

Глава 10. Точное обучение 283

10.1. Изучение функций с конечным носителем	283
10.2. Обучение с использованием функции ReLU	285
10.2.1. Представление максимумов.....	285
10.2.2. Ширина и глубина	289
10.3. Теорема Колмогорова–Арнольда–Спречера.....	292
10.4. Интегральная формула Ириэ и Мияке.....	294
10.5. Получение точного представления в результате обучения не всегда возможно	296
10.6. Континуальное количество нейронов	296
10.7. Аппроксимация вырожденными ядрами.....	301
10.8. Примеры вырожденных ядер.....	303
10.9. Глубокие сети.....	305
10.10. В заключение	306
10.11. Упражнения	307

ЧАСТЬ III. ОБРАБОТКА ИНФОРМАЦИИ..... 309

Глава 11. Представление информации 310

11.1. Информационное наполнение	310
11.2. Цели, пригодные для изучения	313

11.3. Потеря информации	316
11.4. Восстановимая информация	318
11.5. Примеры представления информации	321
11.5.1. Информация об индикаторных функциях	321
11.5.2. Классический перцептрон	323
11.5.3. Линейный нейрон	323
11.5.4. Сигмоидный нейрон	324
11.5.5. Арктангенсный нейрон	325
11.5.6. Нейронная сеть классических перцептронов	325
11.5.7. Треугольник как объединение прямоугольников	326
11.5.8. Сеть сигмоидных нейронов	327
11.5.9. Дополнительные замечания	329
11.6. Сжимаемые слои	329
11.7. Сжимаемость слоев	332
11.8. Функциональная независимость	333
11.9. В заключение	335
11.10. Упражнения	335

Глава 12. Оценка информационной емкости 339

12.1. Энтропия и свойства	339
12.1.1. Энтропия случайной переменной	339
12.1.2. Энтропия при смене координат	340
12.2. Поток энтропии	343
12.3. Энтропия входного слоя	345
12.4. Линейный нейрон	347
12.5. Автокодировщик	348
12.6. Условная энтропия	349
12.7. Взаимная информация	351
12.8. Приложения к глубоким нейронным сетям	358
12.8.1. Поток энтропии	359
12.8.2. Зашумленные слои	359
12.8.3. Независимые слои	360
12.8.4. Слои без сжатия	360
12.8.5. Количество признаков	362
12.8.6. Полное сжатие	364
12.9. Емкость сети	365
12.9.1. Типы емкости	366
12.9.2. Входное распределение	367
12.9.3. Выходное распределение	367
12.9.4. Тензор ввода-вывода	368
12.9.5. Матрица ввода-вывода	368
12.9.6. Наличие емкости сети	369
12.9.7. Метод множителей Лагранжа	370
12.9.8. Определение емкости	373
12.9.9. Емкость перцептрона	373
12.10. Информационное сужение	376
12.10.1. Точное формальное решение	377
12.10.2. Информационная плоскость	381

12.11. Обработка информации в наборе данных MNIST	383
12.11.1. Двухслойная сеть прямого распространения.....	384
12.11.2. Трехслойная сеть прямого распространения.....	387
12.11.3. Роль сверточных сетей.....	387
12.12. В заключение	390
12.13. Упражнения	390

ЧАСТЬ IV. ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ 395

Глава 13. Выходные многообразия 396

13.1. Введение в многообразия.....	396
13.1.1. Внутреннее и внешнее.....	400
13.1.2. Касательное пространство	400
13.1.3. Риманова метрика	401
13.1.4. Геодезика	402
13.1.5. Связность Леви–Чивита	403
13.1.6. Подмногообразия.....	404
13.1.7. Вторая фундаментальная форма.....	405
13.1.8. Векторное поле средней кривизны	406
13.2. Связь с нейронными сетями	407
13.3. Пространство параметров.....	408
13.4. Оптимальные значения параметров.....	412
13.5. Метрическая структура.....	416
13.6. Регуляризация	419
13.6.1. Переход к меньшему размеру	420
13.6.2. Регуляризация норм	420
13.6.3. Выбор самого плоского многообразия	421
13.6.4. Усреднение модели	428
13.6.5. Прореживание	430
13.7. В заключение	435
13.8. Упражнения	436

Глава 14. Нейронные многообразия..... 438

14.1. Статистические многообразия.....	438
14.2. Информация Фишера	442
14.3. Нейронное многообразие нейронной сети.....	446
14.4. Метрика Фишера для одного нейрона.....	447
14.5. Матрица Фишера и обратная ей матрица	449
14.6. Метрическая структура Фишера нейронной сети	454
14.7. Натуральный градиент	458
14.8. Алгоритм обучения с натуральным градиентом.....	460
14.9. Логарифмическое правдоподобие и метрика.....	464
14.10. Связь с расстоянием Кульбака–Лейблера.....	465
14.11. Метод имитации отжига.....	469
14.12. В заключение	470
14.13. Упражнения	471

ЧАСТЬ V. ДРУГИЕ АРХИТЕКТУРЫ.....475**Глава 15. Объединение476**

15.1. Аппроксимация непрерывных функций.....	476
15.2. Трансляционная инвариантность.....	479
15.3. Информационный подход.....	480
15.4. Объединение и классификация	482
15.5. В заключение.....	483
15.6. Упражнения	483

Глава 16. Сверточные сети485

16.1. Дискретные одномерные сигналы	485
16.2. Непрерывные одномерные сигналы	487
16.3. Дискретные двумерные сигналы.....	488
16.4. Сверточные слои с одномерными входами	489
16.5. Сверточные слои с двумерным входом.....	492
16.6. Геометрия CNN.....	495
16.7. Эквивариантность и инвариантность.....	495
16.7.1. Группы.....	495
16.7.2. Действия групп на множествах	497
16.7.3. Расширение действий до функций	498
16.7.4. Определение эквивариантности	499
16.7.5. Свертка и эквивариантность	500
16.7.6. Определение инвариантности	502
16.8. В заключение	504
16.9. Упражнения	505

Глава 17. Рекуррентные нейронные сети.....508

17.1. Системы состояний	508
17.2. RNN.....	510
17.3. Информация в RNN	511
17.4. Функции потерь.....	513
17.5. Обратное распространение ошибки во времени.....	514
17.6. Проблемы градиента.....	517
17.7. Ячейки LSTM	518
17.8. Глубокие RNN.....	519
17.9. В заключение	520
17.10. Упражнения.....	520

Глава 18. Классификация.....523

18.1. Отношения эквивалентности	523
18.2. Энтропия разбиения.....	525
18.3. Решающие функции.....	526
18.4. Карты решений с векторами прямого кодирования	528
18.5. Линейная разделимость	531
18.6. Выпуклая разделимость	535

18.7. Сжатие к центру.....	537
18.8. Обучение карт решений	537
18.8.1. Линейные карты решений.....	537
18.8.2. Нелинейные карты решений.....	544
18.9. Границы принятия решений.....	546
18.10. В заключение	547
18.11. Упражнения	547
Глава 19. Генеративные модели	549
19.1. Необходимость генеративных моделей	549
19.2. Оценка плотности	550
19.3. Состязательные игры	553
19.4. Генеративно-состязательные сети.....	555
19.5. Генеративные сети согласования моментов	563
19.6. В заключение	564
19.7. Упражнения	564
Глава 20. Стохастические сети	567
20.1. Стохастические нейроны	567
20.2. Распределение Больцмана	570
20.3. Машины Больцмана.....	573
20.4. Алгоритм обучения Больцмана.....	576
20.5. Вычисление распределения Больцмана	578
20.6. Энтропия распределения Больцмана	580
20.7. Информационная метрика Фишера	581
20.8. Применение машин Больцмана.....	583
20.9. В заключение	587
20.10. Упражнения	587
Советы и решения	590
Глава 1.....	590
Глава 2.....	592
Глава 3.....	594
Глава 4.....	596
Глава 5.....	599
Глава 6.....	601
Глава 7.....	604
Глава 8.....	606
Глава 9.....	608
Глава 10.....	610
Глава 11.....	611
Глава 12.....	614
Глава 13.....	618
Глава 14.....	621
Глава 15.....	626
Глава 16.....	628
Глава 17.....	630

Глава 18.....	631
Глава 19.....	633
Глава 20.....	635
ПРИЛОЖЕНИЯ	641
Приложение А. Теория множеств	642
Приложение В. Тензоры	644
Приложение С. Теория меры.....	645
С.1. Информация и $\mathfrak{S}$ -алгебры	645
С.2. Измеримые функции	647
С.3. Меры	648
С.4. Интегрирование в мере.....	650
С.5. Меры образов.....	651
С.6. Неопределенные интегралы	652
С.7. Теорема Радона–Никодима	652
С.8. Теоремы Егорова и Лузина	653
С.9. Меры со знаком	653
Приложение D. Теория вероятности	656
D.1. Общие определения	656
D.2. Примеры	656
D.3. Ожидание	657
D.4. Дисперсия	657
D.5. Информация, генерируемая случайными величинами	659
D.5.1. Фильтрация.....	660
D.5.2. Условные ожидания.....	660
D.6. Типы сходимости.....	661
D.6.1. Сходимость по вероятности	661
D.6.2. Сходимость почти наверняка	661
D.6.3. L^p -сходимость	662
D.6.4. Слабая сходимость.....	662
D.7. Функция логарифмического правдоподобия	663
D.8. Броуновское движение.....	665
Приложение E. Функциональный анализ	666
E.1. Банаховы пространства	666
E.2. Линейные операторы	667
E.3. Теорема Хана–Банаха.....	667
E.4. Гильбертовы пространства	668
E.5. Теоремы о представлении.....	669
E.6. Теорема о фиксированной точке	671
Приложение F. Вещественный анализ	673
F.1. Теорема об обратной функции	673

F.2. Дифференциация в обобщенном смысле674

F.3. Сходимость последовательностей функций675

Приложение G. Линейная алгебра676

G.1. Собственные значения, норма и обратная матрица.....676

G.2. Псевдоинверсия Мура–Пенроуза679

Библиография683

Предметный указатель691

Предисловие 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.

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 – возможно, ошибку в основном тексте или программном коде, – мы будем очень благодарны, если вы сообщите нам о ней. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Если вы найдете какие-либо ошибки в коде, пожалуйста, сообщите о них главному редактору по адресу dmkpress@gmail.com, и мы исправим это в следующих тиражах.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится 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 dmkpress@gmail.com.

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

Предисловие

Многочисленные коммерческие применения нейронных сетей показали их экономическую эффективность. Нейронные сети внедряются в новые технологии, которые в настоящее время успешно используются ведущими компаниями, такими как Google, Microsoft, Facebook, IBM, Apple, Adobe, Netflix, NVIDIA и Baidu.

Нейронная сеть – это совокупность вычислительных блоков, связанных между собой и называемых нейронами. Каждый такой блок выдает действительный результат, называемый активацией. Входные нейроны активируются датчиками, воспринимающими состояние окружающей среды, а другие нейроны активируются результатами, выдаваемыми предыдущими нейронами. Такая организация позволяет нейронам обмениваться сообщениями, усиливать связи, ведущие к успешному решению задачи, и ослаблять ведущие к неудаче.

Эта книга описывает, как работают нейронные сети, с математической точки зрения и предлагает оценивать успех нейронных сетей не методом проб и ошибок, а путем четкого математического анализа. Основная цель настоящей работы – представить на точном языке современной математики идеи и концепции нейронных сетей, которые ныне понимаются преимущественно интуитивно. Книга сочетает старые добрые концепции классической математики и современные идеи глубокого обучения. Основное внимание уделяется математической стороне, потому что в современной тенденции развития многие математические аспекты умалчиваются, и в большинстве статей подчеркиваются только детали, касающиеся информатики и практического применения.

Делая акцент на математике, лежащей в основе нейронных сетей, книга также устанавливает связь с лагранжевой механикой, функциональным анализом, численными методами, дифференциальной геометрией, теорией групп и физикой.

Книга не задумывалась ни как исчерпывающее руководство по нейронным сетям и методам глубокого обучения, ни как сборник рецептов машинного обучения. Тем не менее она подробно описывает основные математические принципы, которые управляют этой отраслью науки, и дает четкое теоретическое обоснование, почему нейронные сети хорошо работают во многих приложениях, предоставляя студентам инструменты, необходимые для аналитического понимания этой области. Эта книга не затрагивает вопросов программирования, и вы не найдете в ней примеров кода. Большинство компьютерных программ, реализующих нейронные сети, быстро устаревают, поэтому книга рассматривает математические концепции, образующие фундамент алгоритмов нейронных сетей, без примеров их реализации на каком-либо конкретном языке программирования.

Читателям, интересующимся практическими аспектами нейронных сетей, включая примеры их реализации, я предлагаю обратиться к вышедшим недавно книгам по этой теме, демонстрирующим реализации алгоритмов машинного обучения на различных языках программирования и с применением различ-

ных фреймворков, таких как TensorFlow, Python или R. Вот несколько примеров таких книг: Ф. Шолле «Deep Learning with Python»¹, К. Элбон «Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning»², С. Рашка и В. Мирджалили «Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow»³, Т. Бейсолоу «Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and Implementing Deep Learning Models Using R», Дж. Паттерсон и А. Гибсон «Deep Learning: A Practitioners Approach»⁴, О. Жерон «Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems»⁵ и Ф. Дюваль «Deep Learning for beginners: A Practical Guide with Python and TensorFlow».

Эта книга адресована аспирантам, изучающим математику, электротехнику, статистику, теоретическую информатику и эконометрику, которые хотели бы познакомиться с теоретическими основами работы нейронных сетей. В этом смысле книга больше полезна исследователям, чем практикам. Для большей полноты в книгу добавлено несколько приложений со сведениями по теории меры, теории вероятностей и линейной алгебре, а также по теории функций действительной переменной. В зависимости от имеющихся знаний разные читатели найдут полезными разные приложения. Например, большинство аспирантов, изучающих математику, которые обычно знакомы с теорией функций действительной переменной, сочтут большую часть книги вполне доступной, в то время как студентам-статистикам, изучающим минимальный объем теории мер и вероятностей, будет полезно прочитать аналитическую часть приложения. Наконец, студентам-информатикам, которые обычно не получают глубоких математических познаний, большая часть приложений окажется полезной. Минимальные знания, которыми необходимо обладать для чтения этой книги: основы математического анализа, линейная алгебра с векторами и матрицами и элементы теории вероятностей.

Однако эта книга рекомендуется студентам, уже прошедшим вводный курс по машинному обучению и заинтересованным в дальнейшем углублении их понимания машинного обучения с математической точки зрения. Книгу можно использовать как учебник для изучения в аспирантуре в течение одного или двух семестров с охватом частей I–II и I–V соответственно.

Книга содержит изрядное количество упражнений, в основном решенных или содержащих подсказки. Вам встретятся и очень простые, и очень сложные упражнения, и многие из них являются приложениями или расширениями

¹ Шолле Ф. Глубокое обучение на Python. Питер, 2022. ISBN 978-5-4461-0770-4. – Прим. перев.

² Элбон К. Машинное обучение с использованием Python. Сборник рецептов. БХВ, 2019. ISBN: 978-5-9775-4056-8. – Прим. перев.

³ Рашка С., Мирджалили В. Python и машинное обучение. Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2. Диалектика-Вильямс, 2021. ISBN: 978-5-907203-57-0. – Прим. перев.

⁴ Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика. ДМК Пресс, 2018. ISBN: 978-5-97060-481-6. – Прим. перев.

⁵ Жерон О. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. Концепции, инструменты и техники для создания интеллектуальных систем. Диалектика, 2020. ISBN: 978-5-907203-33-4. – Прим. перев.

представленных результатов. Список ссылок на литературу не является исчерпывающим, потому что ни о какой полноте в принципе не может идти и речи.

Несмотря на большое количество книг по нейронным сетям, опубликованных за последние несколько лет, очень немногие из них носят теоретический характер и обсуждают «математические концепции», лежащие в основе нейронных сетей. Одна из известных таких книг, содержащая всесторонний обзор литературы по глубокому обучению, – это книга Гудфеллоу и его коллег [46], которую можно использовать как в качестве учебника, так и в качестве справочного пособия. Еще одна весьма интересная работа – онлайн-книга Нильсена [92]. В более старой книге Роя [101] вы найдете классическое введение в теорию нейронных сетей. Хочу еще отметить книгу Хастая и его коллег [53], которая содержит обсуждение статистической природы алгоритмов обучения, таких как «искусственные нейронные сети». Математические концепции, лежащие в основе нейронных сетей, также обсуждаются в книгах Мори и его коллег [88] и Шалев-Шварца и его коллег [110].

Даже притом что предыдущие книги охватывают важные аспекты, связанные со статистическими основами или вычислительной сложностью глубокого обучения, все еще наблюдается нехватка литературы в области, которую пытается охватить эта книга. Здесь предпринята попытка предоставить полезный вводный материал и обсудить типы функций, которые могут быть представлены нейронными сетями глубокого обучения.

ОБЗОР КНИГИ

Книга состоит из четырех основных частей, расположенных по возрастанию уровня сложности обсуждаемых тем. В первой части дается доступное введение в теорию нейронных сетей; во второй части нейронные сети рассматриваются как универсальные аппроксиматоры; в третьей – нейронные сети исследуются как процессоры информации. Последняя часть посвящена теории, лежащей в основе нескольких интересных тем, таких как объединение, свертка, сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN), генеративно-сопоставительные нейронные сети (GAN), машины Больцмана и классификация. Для большей полноты книга содержит приложение, в котором представлены такие темы, как теория меры, теория вероятностей, линейная алгебра и функциональный анализ. Каждая глава заканчивается подведением итогов и набором упражнений. Решения и подсказки к предлагаемым упражнениям даются в конце книги.

Часть I

Это самая простая часть книги. Она содержит введение в такие классические темы, как функции активации, функции стоимости и алгоритмы обучения в нейронных сетях.

В главе 1 представлены некоторые повседневные задачи, приведшие к появлению понятия абстрактного нейрона. Здесь читатель познакомится с процессом регулирования скорости потока жидкости или силы электрического тока, питающего резервуар, аккумулятор, транзистор и т. д., запускающим

определенную функцию активации. Оптимизация процесса включает в себя минимизацию функции стоимости (затрат), такой как объем, энергия, потенциал и т. д. Большое количество нейронов (нейронных модулей) могут работать вместе как нейронная сеть. В качестве примера в этой главе приводится связь между линейной регрессией и нейронными сетями.

Чтобы изучить нелинейную целевую функцию, нейронная сеть должна использовать нелинейные функции активации. Выбор этих функций активации определяет тип сети. Глава 2 предлагает систематическое представление набора функций активации, которые можно найти в литературе. Они подразделяются на три основных класса: сигмоидные (логистическая, гиперболический тангенс, *softsign*, арктангенс), Г-образные (ReLU, PReLU, ELU, SELU) и выпуклые (гауссова, двойная экспоненциальная).

В процессе обучения нейронной сети приходится корректировать параметры, чтобы добиться минимизации определенной целевой функции. Эта функция также известна под названиями: функция стоимости, функция затрат, функция ошибок или функция потерь. В главе 3 описаны некоторые наиболее известные функции стоимости, используемые в нейронных сетях. К ним относятся: функция верхней границы ошибки, функция L^2 -ошибки, функция среднеквадратической ошибки, перекрестная энтропия, расстояние Кульбака–Лейблера, расстояние Хеллингера и другие. Некоторые из этих функций стоимости лучше подходят для изучения случайных величин, а другие – для изучения детерминированных функций.

В главе 4 представлена серия классических алгоритмов минимизации. Они необходимы для минимизации соответствующей функции стоимости. Чаще других используется алгоритм градиентного спуска, который будет представлен во всех подробностях. Также в этой главе будут представлены такие алгоритмы, как метод линейного поиска, метод моментов, имитация отжига, AdaGrad, Adam, AdaMax, гессиан и методы Ньютона.

Глава 5 знакомит с понятием абстрактного нейрона и представляет несколько классических типов нейронов, такие как перцептрон, сигмоидный нейрон, линейный нейрон и нейрон с континуальным входом. Здесь же вы найдете некоторые приложения логистической регрессии и классификации.

Глава 6 завершает знакомство с нейронными сетями. Здесь подробно рассматриваются архитектуры сетей и метод обратного распространения ошибки, используемый для обучения сети.

Часть II

Основная идея этой части: нейронные сети являются универсальным аппроксиматорами, т. е. механизмами, выходные данные которых могут аппроксимировать большое количество типов целей, таких как непрерывная, квадратично интегрируемая или линейно интегрируемая функция, а также измеримые функции.

Глава 7 знакомит с рядом классических аппроксимаций теорем аналитического толка. Этот мощный инструмент с приложениями для обучения содержит теорему Дини, теорему Арцела–Асколи, теорему Стоуна–Вейерштрасса, тауберovu теорему Винера и принцип сжатия.

В главе 8 рассматривается случай, когда входная переменная ограничена и одномерна. Помимо простоты, этот случай позволяет рассмотреть обучение на элементарном уровне и носит конструктивный характер. Здесь рассматриваются два случая нейронных сетей: со множеством перцептронов и сигмоидных нейронов с одним скрытым слоем. Еще два раздела посвящены обучению с помощью функций ReLU и Softplus.

Глава 9 отвечает на вопрос, какие функции можно изучить с помощью нейронных сетей с одним скрытым слоем. Он основан на теории аппроксимации результатов, разработанной Фунахаши, Хорником, Стинчкомбом, Уайтом и Цыбенко в конце 1980-х – начале 1990-х годов. В главе приводятся математические доказательства того, что нейронные сети с одним скрытым слоем могут изучать непрерывные, L^1 - и L^2 -интегрируемые, а также измеримые функции на компактном множестве.

Глава 10 посвящена случаю точного обучения, а именно случаю, когда сеть может точно воспроизвести желаемую целевую функцию. Глава содержит утверждения, касающиеся точного изучения функций с конечным носителем, функции максимума и кусочно-линейной функции. Здесь также будут представлены теорема Колмогорова–Арнольда–Шпречера, интегральная формула Ириэ и Мияке, а также точное обучение с использованием интегральных ядер при применении большого количества нейронов.

Часть III

Главная идея этой части: нейронные сети можно рассматривать как информационные процессоры. Входные данные нейронной сети содержат информацию, которая обрабатывается множеством скрытых слоев сети до тех пор, пока не будут выбраны характеристические признаки. Выходной слой содержит сжатую информацию, используемую для определенных целей, например для классификации. Работу нейронной сети можно сравнить с тем, что делает скульптор в процессе изготовления скульптуры. На первом этапе скульптор готовит мраморный блок, а потом высекает грубые формы. Затем он продолжает обозначать общие особенности и заканчивает тонкими деталями. Подобно скульптору, отсекающему лишнее на каждом шаге, нейросеть отсекает в каждом слое определенное количество ненужной информации, так что в конце остаются только нужные признаки.

Глава 11 моделирует информацию, обрабатываемую нейронной сетью, с точки зрения сигма-полей. В каждом слое прямого распространения нейронной сети теряется какая-то часть информации. Здесь даются необходимые условия для слоев с тривиальными потерями информации, а также приводится информационное описание нескольких типов нейронов. Например, перцептрон выявляет полуплоскости в предоставленной информации, а мультиперцептрон способен выявлять более сложные структуры. Сжимаемые слои изучаются также с точки зрения сжатия информации.

Глава 12 является логическим продолжением предыдущей. В ней пойдет речь о количественной оценке распространения потока информации через слои нейронной сети. Для этого используются такие инструменты, как энтропия, условная энтропия и взаимная информация. В качестве приложения

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru