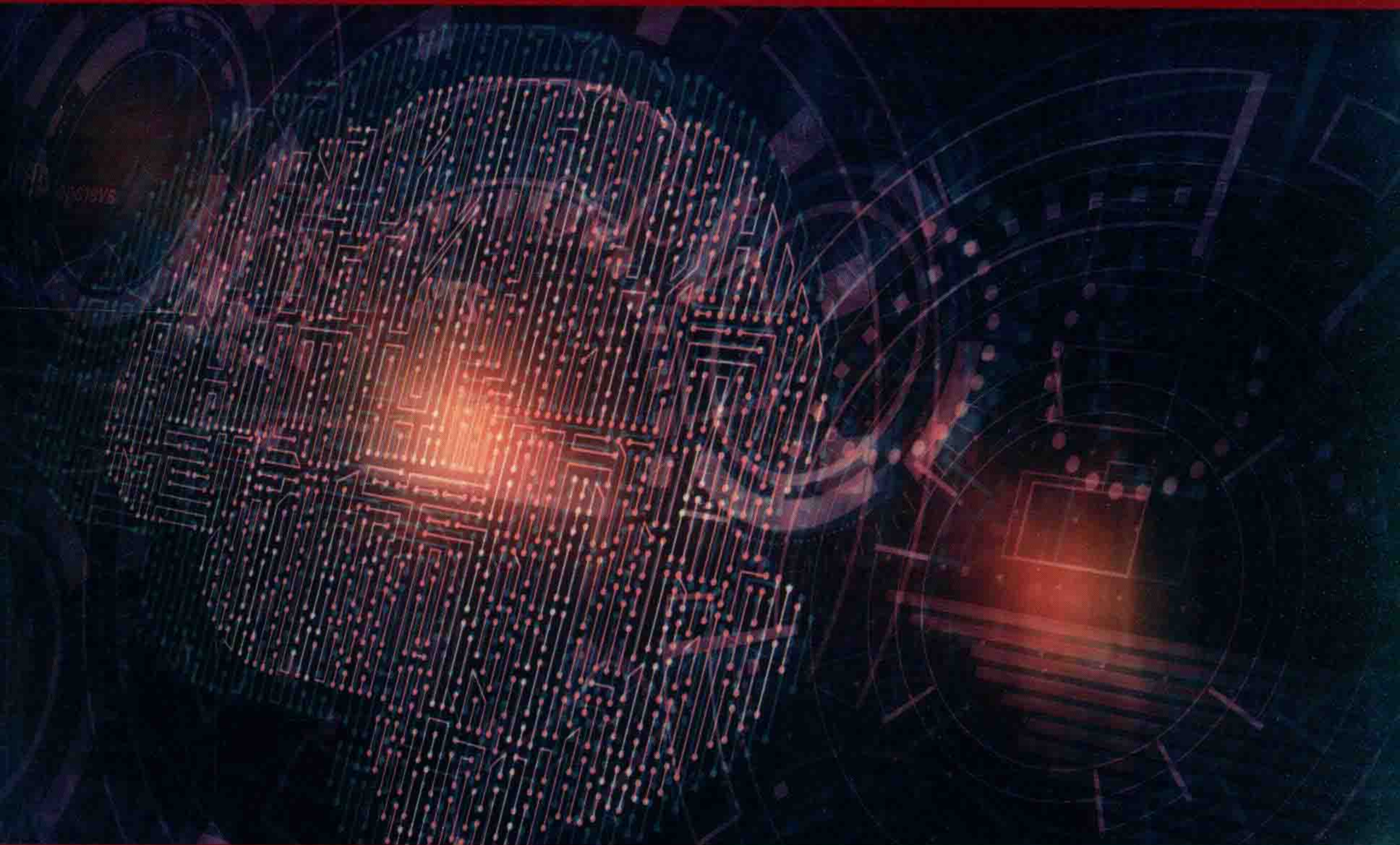


深度学习实践指南

—— 基于R语言

[英] 尼格尔·刘易斯 (N.D. Lewis) 著 沙瀛译



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

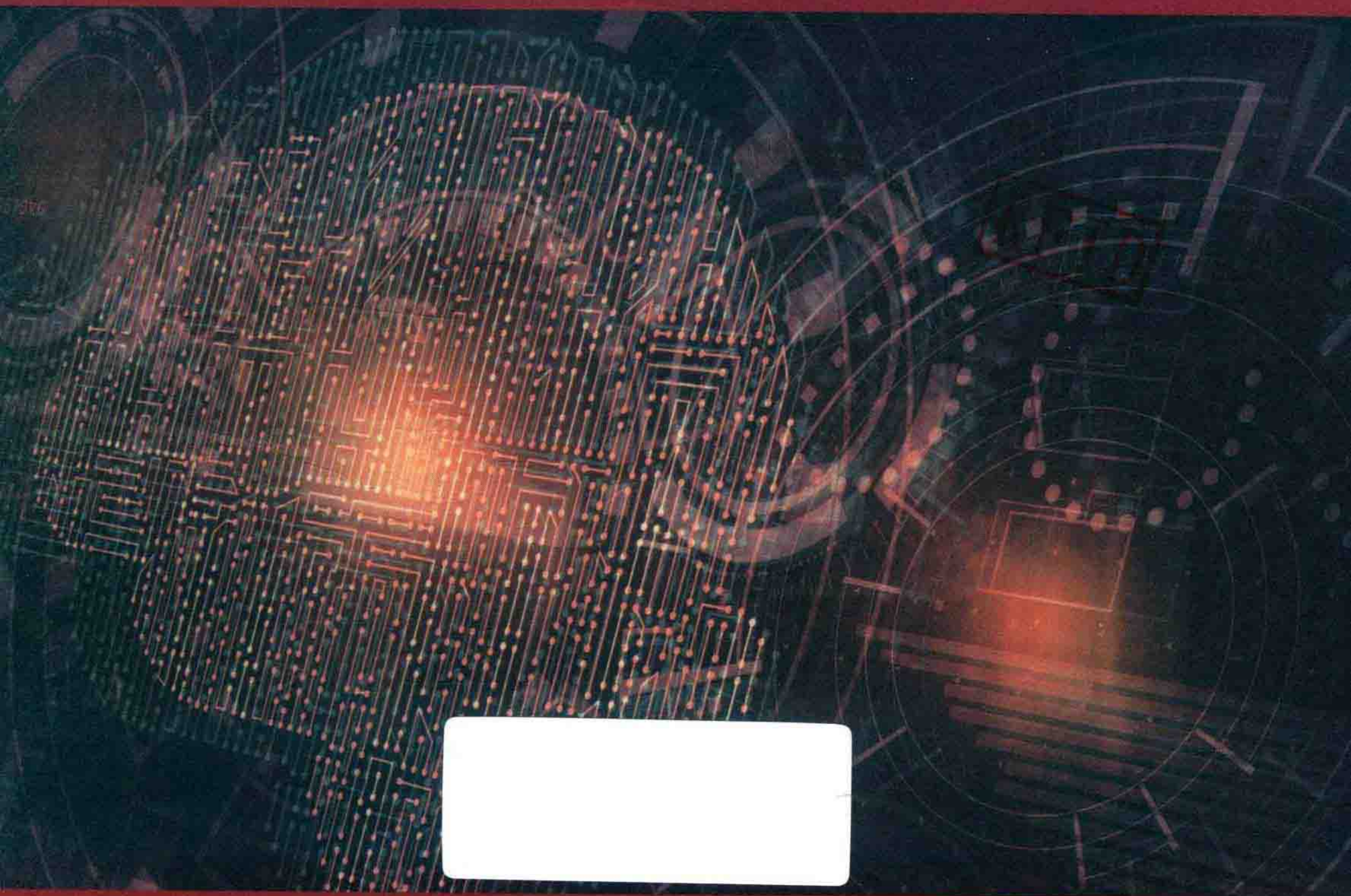


深度学习系列
DEEP LEARNING SERIES

异步图书
WWW.APUBIT.COM.CN

深度学习实践指南 —— 基于R语言

[英] 尼格尔·刘易斯 (N.D. Lewis) 著 沙瀛译



人民邮电出版社
北京

图书在版编目 (C I P) 数据

深度学习实践指南：基于R语言 / (英) 尼格尔·刘易斯 (N. D. Lewis) 著；沙瀛译. — 北京：人民邮电出版社，2018.6

(深度学习系列)

ISBN 978-7-115-47777-4

I. ①深… II. ①尼… ②沙… III. ①程序语言—程序设计 IV. ①TP312

中国版本图书馆CIP数据核字(2018)第045187号

版权声明

Simplified Chinese translation copyright ©2017 by Posts and Telecommunications Press ALL RIGHTS RESERVED

Deep Learning Made Easy with R:A Gentle introduction For Data Science, by Dr. N. D Lewis ISBN 9781519514219

Copyright © 2016 by Dr. N. D Lewis

本书中文简体版由 Dr. N. D Lewis 授权人民邮电出版社出版。未经出版者书面许可，对本书的任何部分不得以任何方式或任何手段复制和传播。

版权所有，侵权必究。

-
- ◆ 著 [英] 尼格尔·刘易斯 (N.D. Lewis)
 - 译 沙 瀛
 - 责任编辑 陈冀康
 - 执行编辑 武晓燕
 - 责任印制 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京鑫丰华彩印有限公司印刷
 - ◆ 开本：720×960 1/16
 - 印张：12.5
 - 字数：156 千字 2018 年 6 月第 1 版
 - 印数：1—2 500 册 2018 年 6 月北京第 1 次印刷

著作权合同登记号 图字：01-2016-1424 号

定价：69.00 元

读者服务热线：(010)81055410 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字 20170147 号

内容提要

深度学习的概念源于人工神经网络的研究，它是机器学习中一种基于对数据进行表征学习的方法。

本书是一本详细的、实用的深度学习实践指南。它共有 8 章，详细讲解了深度神经网络、Elman 神经网络、Jordan 神经网络、自编码器、堆叠自编码器、限制玻尔兹曼机的相关知识。本书并没有详细介绍那些深奥的数字公式，它旨在解释深度学习模型是如何工作的，让读者学会如何构建最成功的深度学习模型，并将其用于数据挖掘，从而让读者迅速地学以致用，可以用深度学习构建更智能的应用。

本书适合数据科学家、各领域的研究人员阅读，也适合其他对深度学习感兴趣的人士阅读。

致谢

特别感谢：

我的妻子安吉拉，感谢她的耐心和持之以恒的鼓励；

我的女儿迪安娜，感谢她为了这本书和我的网站拍摄了几百幅照片；

我的以往著作的读者，感谢你们提出的问题和建议。

序

本书适用于任何对深度学习感兴趣的人。深度学习并不只是对跨国公司或大学的研究部门有意义。很多人都可以将将在本书中学到的深度学习的思想和实用技术应用到各自的领域中去，这些人可以是小广告公司工作的数据科学家、受雇于本地宠物食品公司的3个决策科学家组成的小团队、完成深度学习大作业的学生、承担地方健康权威机构预测项目的独立咨询师等。通过本书了解深度学习并从中受益并不要求你必须是个天才的统计学家或编码高手。

本书的目的是向数据科学家、各领域的研究人员、业余爱好者或者其他对深度学习感兴趣的人士介绍深度学习技术。本书并不准备从全面综合的理论角度来阐述这项技术。本书的意图是尽可能地隐藏技术细节，但是仍然能向读者清晰地传达使用 R 语言的深度学习技术可以做什么以及怎么做。当然重点是在“怎么做”，因为只有了解到怎么做，才会驱使你在面对实际问题的时候提出创新性的实用解决方案。正如 Anne Isabella Richie^①，作家 Willam Makepeace Thackeray^② 的女儿，在其小说《戴蒙德夫人》（Mrs. Dymond）所说的那样“授人以鱼不如授人以渔”。

本书写作的主要目的是想将强有力的深度学习技术应用到日常实践

① Anne Isabella Richie (1837.6.9 – 1919.2.26)，英国作家，萨克雷先生的大女儿。她是维多利亚时代晚期文坛的中心人物，被认为是她父亲的文学遗产的最好的看守人。她的短篇小说将传统童话故事融入维多利亚时代的环境。1885 年，她的小说《戴蒙德夫人》包含了被认为是最早在英文中使用的谚语“授人以鱼不如授人以渔”。

② Willam Makepeace Thackeray (1811.7.18 – 1863. 12.24)，英国作家，其代表作品是世界名著《名利场》。与狄更斯齐名，为维多利亚时代的代表小说家。还著有《班迪尼斯》等作品。

中。因此本书的内容主要面向需要进行数据分析和建模的相关人员。本书主要介绍那些已经被证明行之有效而且可以在短时间内迅速消化吸收然后应用的深度学习的相关技术、思想和策略。

在很多场合，笔者都会被各种专业背景的人士问道：“如何能够迅速理解深度学习技术然后将其应用到本专业领域？”这个问题的答案往往包括阅读复杂的数学教材，然后用各种计算机语言，如 C、C++、Java 等编程实现复杂的数学公式。

随着 R 语言的出现，使用深度学习技术变得比以往任何时候都简单。本书就是提供这种快速入门的实践指南。本书将向你一步步地展示如何利用 R 语言统计工具包构建各种深度学习模型。描述清晰的应用示例可以直接在 R 语言环境中运行。

对于实践者来说最枯燥的部分就是算法的相关理论部分。尽管这个深度学习主题中最“恐怖”的部分将是理论专家所感兴趣的，但是对于实践者来说却是最不重要的部分，通过使用 R 工具包几乎可以将其从实践者的必备知识中完全删除。当然不可避免地会有少数必备的知识点，本书会充分讨论这些知识点。然而这本书是实践指南，所以“卷起袖子，开始干吧”，我不会花费大量的时间拘泥于算法的理论证明、推理等。

处理实际数据的研究人员主要关注如何在数据中提取有意义的特征和结构。因此先研究下其他研究人员是如何应用这项技术，这通常是一个很好的想法。这主要是因为实践往往与从书本里、教室里学到的知识有显著的差异。以实践为目标，并且为了加快学习过程，本书会有大量的深度学习的应用实例。

应用实例覆盖了各个领域，并辅以实际行动研究，这会帮助读者一步步学会如何使用 R 语言构建深度学习模型。在每章结尾的注释小节中会给出进一步自学的详细的参考文献。因为 R 语言符合当前的时代精神（开源），本书所引用的文献绝大多数都可以免费获得。

R 语言的新手也可以很轻松地使用这本书，不需要任何先验知识。通过本书学习 R 语言的最佳实践就是敲入实例代码，然后阅读注释就可以了。R 语言及入门指南可以访问 R 官网。有很多资料都十分适合于 R 语言的初学者。

深度学习不仅仅是理论证明和数学推导，它最终是关于真实的现实世界、真实的人类，最终是使用机器学习算法解决现实生活中的实际问题。不论你是谁、来自哪里，也不论你的教育背景如何，都可以使用本书介绍的思想和方法。只要使用合适的工具、小小的坚持和正确的指导，我个人坚信只要是对深度学习感兴趣的人都可以通过本书的学习成功地学会应用此项技术。

著名画家 Vincent 曾经说过“伟大的事业并不是一蹴而就的，而是由一系列微小成果累积而成的。”这本入门指南将提供详细的、实用的、战术层面的指导，手把手教会你如果构建和使用深度学习模型。这本书是我在深度学习领域前期工作的扩展、修订、更新的版本，其中浓缩了我的最佳实践思路。

我发现，如果能够面对大量的各种模型、工具和应用，其效果将远远超过只聚焦在某些按照特定原则选择的模型、工具和应用。了解如何构建和使用深度学习模型无疑会大大充实你的个人“工具箱”。

古希腊哲学家 Epicurus^① 说过“我写下这个不是为了泛泛之终生，只是为了你；每个人都足以成为另外一个人的观众”。尽管本书中的思想可以触及成千上万的人，但是在撰写本书的时候，我时刻牢记 Epicurus 提出的原则，书中每页的内容都是为了每一个独特的你而写的。

^① Epicurus（希腊文：Ἐπίκουρος，公元前 341～公元前 270 年），古希腊哲学家、无神论者（被认为是西方第一个无神论哲学家），伊壁鸠鲁学派的创始人。他的学说的主要宗旨就是要达到不受干扰的宁静状态，并要学会享乐。

前言

这是一本让你直接“卷起袖子”马上动手测试数据、进行R语言编码的书。输入示例代码、阅读注释和参考文献、对数据多做实验会让你获得最大收益。测试大量的应用示例并阅读注释和参考文献可以扩大你的知识面，加深你对深度学习的本质理解，增强你的实践技能。

使用本书的方法至少还有另外两种。你可以把它当作参考书。可以直接跳到你感兴趣的章节，快速理解R语言是如何实现具体的深度学习技术的。为了获得最大收益，你需要输入本书的示例代码、验证结果，然后根据具体问题对示例代码做相应的调整。另外一种方法是通过快速浏览真实世界的应用实例、说明解释、案例分析、技巧窍门和注意事项等来刺激你产生新的想法。

实践小窍门

如果你使用 Windows 操作系统，可以很轻松地通过 `installr` 包将 R 升级到最新版。输入如下命令：

- `install.packages("installr")`
- `installr::updateR()`

如果你还没有安装本书提及的包，你可以通过输入 `install.packages("包名")` 来下载。例如，为了下载 `RSNNS` 包，你可以在 R 控制窗口输入：


```
install.packages("RSNNS")①
```

一旦包已安装，需要加载后才能使用。在 R 控制窗口输入：

```
require(RSNNS)
```

现在 RSNNS 包就可以使用了。你只需要在开始 R 会话前输入上述命令。

R 中的函数通常有多个参数。在本书的例子中，我主要关注那些快速开发模型所需要的关键参数。在 R 控制窗口输入“? 函数名”就可以获得相应函数参数的相关信息。例如，为了查阅 elman 函数参数的相关信息，需要输入：

```
? elman
```

然后函数的细节和参数的相关信息会显示在你的默认浏览器上。在初步构建了应用领域模型后，强烈建议尝试调整其他的参数看看效果如何。

实践小窍门

你只需要通过 HTTPS 加密连接从 CRAN 下载包即可。直接从合法的 CRAN 镜像网站下载将比从其他网站下载安全保障更高。当用 HTTPS 链接下载包的时候，你可能会遇到类似下面的错误信息：

```
"unable to access index for repository https://cran.rstudio.com/..."
```

这个在 Windows 平台很常见，在版本 R-3.2.2 之前需要激活 internet2 动态链接库。如果你使用的是老版本的 R，在下载新包之前需要输入：

```
setInternet2(TRUE)
```

在贯穿本书的示例代码中已经包含了 set.seed 方法，这样你运行代码所获得的结果就和本书展示的一致了。

R 支持所有的主流操作系统。由于 Windows 平台的受众较多，本书的例子主要使用 Windows 平台下的 R。

^① 在 Windows 平台启动 R 的时候请以管理员身份运行，否则安装包可能会出错。

实践小窍门

想不起来两小时之前你敲了什么命令，不要着急，我也是这样！如果你在同一个 R 会话登录，只需要输入：

```
history(Inf)
```

就会返回当前会话已输入的全部命令。

你不需要把整本书都读完，再在你的专业领域实践深度学习的想法。你几乎可以立刻体验深度学习令人惊讶的威力。你可以直接跳到感兴趣的章节，然后马上测试代码、创建模型、对本领域的数据进行验证分析。

实践小窍门

在 32 位 Windows 平台下，不管机器的实际物理内存是多少，R 只能用到 3G 内存。用下面的命令来验证可以使用多少内存：

```
memory.limit()
```

用下面的命令移除内存中的所有对象：

```
rm(list=ls())
```

正如书名所展示的那样，本书讲解如何理解和构建深度学习模型。更准确地说，本书尝试通过提供一些工具，让你使用 R 更轻松、迅速地构建上述深度学习模型。本书的目标是给读者提供完成上述任务所必需的工具，然后给予必要的解释说明，使读者能够在自己的专业领域独立地完成实际的应用。我希望此过程不仅有益而且有趣。

将本书的思想应用于你的数据集，会获得有益的结果。即使每章你只理解了一个方法，当面临大数据分析环境下的挑战和机遇的时候，也足以有很大的可能创造辉煌。

当你应用这些深度学习模型在本领域获得了成果后，可以给 info@NigelDLewis.com 发邮件让我知晓，我会很高兴收到你的来信。

现在，让我们开始学习之旅吧！

资源与支持

本书由异步社区出品，社区（<https://www.epubit.com/>）为您提供相关资源和后续服务。

配套资源

本书提供如下资源：

- 书中彩图文件；

要获得以上配套资源，请在异步社区本书页面中点击 **配套资源**，跳转到下载界面，按提示进行操作即可。注意：为保证购书读者的权益，该操作会给出相关提示，要求输入提取码进行验证。

提交勘误

作者和编辑尽最大努力来确保书中内容的准确性，但难免会存在疏漏。欢迎您将发现的问题反馈给我们，帮助我们提升图书的质量。

当您发现错误时，请登录异步社区，按书名搜索，进入本书页面，点击“提交勘误”，输入勘误信息，单击“提交”按钮即可。本书的作者和编辑会对您提交的勘误进行审核，确认并接受后，您将获赠异步社区的 100 积分。积分可用于在异步社区兑换优惠券、样书或奖品。

详细信息

写书评

提交勘误

书名:

页码位置 (页码):

勘误内容:

B I U 书 三 的 勘 误

提交

扫码关注本书

扫描下方二维码，您将会在异步社区微信服务号中看到本书信息及相关的服务提示。



与我们联系

我们的联系邮箱是 contact@epubit.com.cn。

如果您对本书有任何疑问或建议，请您发邮件给我们，并请在邮件标题中注明本书书名，以便我们更高效地做出反馈。

如果您有兴趣出版图书、录制教学视频，或者参与图书翻译、技术审校等工作，可以发邮件给我们；有意出版图书的作者也可以到异步社区在线提交投稿（直接访问 www.epubit.com/selfpublish/submission 即可）。

如果您是学校、培训机构或企业，想批量购买本书或异步社区出版的其他图书，也可以发邮件给我们。

如果您在网上发现有针对异步社区出品图书的各种形式的盗版行为，包括对图书全部或部分内容的非授权传播，请您将怀疑有侵权行为的链接发邮件给我们。您的这一举动是对作者权益的保护，也是我们持续为您提供有价值的内容的动力之源。

关于异步社区和异步图书

“异步社区”是人民邮电出版社旗下 IT 专业图书社区，致力于出版精品 IT 技术图书和相关学习产品，为作译者提供优质出版服务。异步社区创办于 2015 年 8 月，提供大量精品 IT 技术图书和电子书，以及高品质技术文章和视频课程。更多详情请访问异步社区官网 <https://www.epubit.com>。

“异步图书”是由异步社区编辑团队策划出版的精品 IT 专业图书的品牌，依托于人民邮电出版社近 30 年的计算机图书出版积累和专业编辑团队，相关图书在封面上印有异步图书的 LOGO。异步图书的出版领域包括软件开发、大数据、AI、测试、前端、网络技术等。



异步社区



微信服务号

目录

第 1 章	简介	1
1.1	什么是深度学习	2
1.2	深度学习解决什么问题	3
1.3	谁在使用深度学习	4
1.4	神经网络基础	6
1.4.1	神经元的角色（作用）	9
1.4.2	激活函数	11
1.4.3	神经网络学习算法	13
1.5	注释	16
第 2 章	深度神经网络	24
2.1	令人惊讶的对深度神经网络的简单剖析	25
2.2	如何用 60 秒或更少的时间解释清楚深度神经网络	26
2.3	3 个绝妙的使用深度神经网络的途径	27
2.3.1	在有雾天气增强能见度	27
2.3.2	让黑客和网络犯罪分子吃些苦头	30
2.3.3	令人难以置信的收缩的图像压缩	31
2.4	如何迅速逼近任意函数	33
2.5	选择多少个神经元	38
2.6	选择最佳神经网络层数的常识	40
2.7	提升深度神经网络性能的 3 种方法	41
2.7.1	用“dropout”方法来提高成功率	41
2.7.2	如何从小批量（mini batching）中获益	44
2.7.3	提前终止（Early stopping）的简单计划	45
2.8	用 R 以难以置信的简单方法来构建深度神经网络	47

2.8.1 构建深度神经网络回归模型的方法	47
2.8.2 聪明人使用深度神经网络回归模型的技巧	51
2.8.3 构建深度神经网络分类器的艺术	54
2.8.4 如何构建多响应变量模型	61
2.9 注释	69

第 3 章 Elman 神经网络 75

3.1 什么是 Elman 神经网络.....	76
3.2 上下文层神经元的作用是什么	77
3.3 如何理解信息的流动	77
3.4 如何用 Elman 神经网络提升效果	78
3.5 使用 Elman 神经网络的 4 种妙招	78
3.5.1 终极天气预报模型	79
3.5.2 如何迅速发现严重的故障	79
3.5.3 提高水质量的创新性想法	80
3.5.4 在股票证券市场如何实现一个“杀手级”应用	80
3.6 构建 Elman 神经网络的简单方法	81
3.7 如何加载工具包	82
3.8 为什么数据可视化是一门科学	82
3.9 转换数据的秘密	85
3.10 如何估计模型.....	88
3.11 创建理想的预测	89
3.12 注释.....	90

第 4 章 Jordan 神经网络 92

4.1 Jordan 神经网络可以解决的 3 个问题	93
4.1.1 风速预测的终极指南	93
4.1.2 如何对蛋白质的相互作用分类	93
4.1.3 深度学习在西班牙语方面的应用	94
4.2 R 语言 Jordan 神经网络模型的基本要素	94
4.3 寻找合适的包	95
4.4 转换数据的方法	96
4.5 如何选择训练样本	98

4.6	用这个技巧来预估你的模型	98
4.7	注释	100
第 5 章	自编码器的秘密	102
5.1	绝地控心术	103
5.2	秘密揭晓	104
5.3	可以直接检验的实用定义	106
5.4	如何拯救巴西热带草原赛拉多 (Cerrado)	106
5.5	需要了解的基本要素	107
5.6	稀疏自编码器的强大益处	108
5.7	理解 Kullback-Leibler 距离	108
5.8	对稀疏自编码器的 3 个永恒的教训	109
5.9	好莱坞、生物统计学和稀疏自编码器的混合	109
5.10	如何利用 R 语言快速使用自编码器	111
5.11	在你自己的数据科学项目使用 R 语言	116
5.12	注释	123
第 6 章	堆叠自编码器简介	125
6.1	深度学习大师的秘密武器	126
6.2	最佳睡眠时间	127
6.3	不超过 5 分钟就可以构建一个堆叠自编码器	130
6.4	什么是去噪自编码器	131
6.5	随机“调味剂”	132
6.6	去噪自编码器的两个核心任务	133
6.7	如何理解堆叠去噪自编码器	133
6.8	一个惊人的实际应用	134
6.8.1	一个创新想法	136
6.8.2	Chen、Li、Yang 是如何训练他们的模型的	138
6.8.3	如何避免塞壬的歌声	138
6.8.4	作者给读者提出的一个挑战	140
6.9	用 R 语言构建去噪自编码器的捷径	141
6.10	注释	146