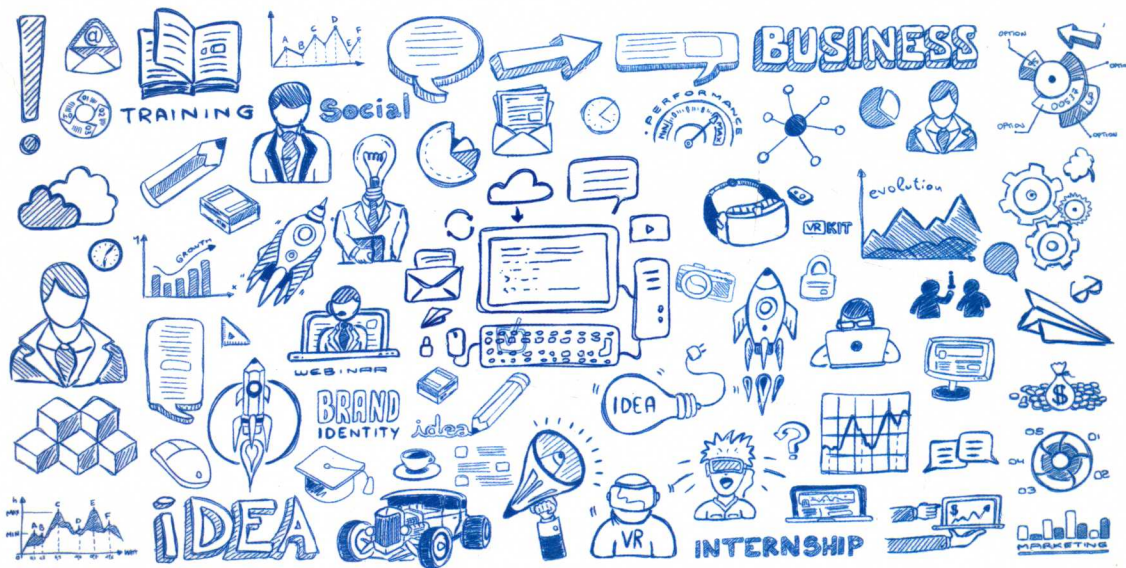


[英] 尼格尔·刘易斯 (N.D.Lewis) 著 邓世超 译



[英] 尼格尔·刘易斯 (N.D.Lewis) 著 邓世超 译



图书在版编目 (C I P) 数据

深度学习企业实战：基于R语言 / (英) 尼格尔·刘易斯 (N. D. Lewis) 著；邓世超译. — 北京：人民邮电出版社，2019.6
ISBN 978-7-115-51009-9

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

中国版本图书馆CIP数据核字(2019)第053187号

版权声明

Simplified Chinese translation copyright ©2019 by Posts and Telecommunications Press.
ALL RIGHTS RESERVED
Deep Learning For Business With R, ISBN-13: 978-1537075044
by N.D. Lewis
Copyright © 2016 by N.D. Lewis

本书中文简体版由作者 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 邓世超
 - 责任编辑 胡俊英
 - 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>
 - 固安县铭成印刷有限公司印刷
 - ◆ 开本：800×1000 1/16
 - 印张：13
 - 字数：249千字 2019年6月第1版
 - 印数：1—2000册 2019年6月河北第1次印刷
 - 著作权合同登记号 图字：01-2016-9392号
-

定价：59.00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

深度学习是机器学习的一个分支，它能够使计算机通过层次概念来学习经验和理解世界。同时，深度学习也是非常贴近 AI 的一个技术分支，得到了越来越多人的关注。本书侧重于 R 语言与深度学习的结合，旨在通过通俗易懂的语言和实用技巧的介绍，帮助读者了解深度学习在商业领域的应用。

本书包含 12 章，涉及基本的 R 编程技巧和深度学习原理，同时介绍了神经网络和深度学习在商业分析中的应用。除此之外，本书还介绍了神经网络的学习机制、激活函数等内容，并且给出了新闻分类、客户维系方法、消费预测、产品需求预测等实用策略。

本书注重实用性，不对读者做过多的技术要求，适合所有想通过 R 编程来了解深度学习，并对其商业化应用感兴趣的读者。

致谢

我特别要感谢：

我的妻子 Angela，感谢她的耐心和不断鼓励；

我的女儿 Deanna 和 Naomi，感谢她们为本书和我的网站拍摄了数百张照片；

本书的早期读者，感谢他们为我提供的宝贵建议。

序言

掌握深度学习并进行商业分析

深度学习就像拥有魔法一样在媒体上引起了轰动。但你该如何入门呢？本书可以带你完成一次轻松而有趣的旅程，在 R 中通过深度神经网络构建自己的业务模型。本书采用平实的语言，在 R 语言中提供了直观、实用、非数学的，且易于遵循的指南来传达最佳的理念，既包含出色的技术，又提供了可用的解决方案。

本书对读者无经验要求，假定读者从来没有学过线性代数，更不喜欢看到导数之类的东西，也不喜欢复杂的计算机代码，读者阅读本书的目的是希望通过通俗易懂的语言，了解如何使用深度神经网络以解决商业问题，并希望亲自动手实践一下。

本书面向有以下需求的读者：

- 侧重语言的解释说明，避免枯燥的数学推导；
- 拥有真实数据的商业应用程序；

- 对于 R 中的运行实例，可以轻松地理解并实践。

快速上手：深度学习不仅仅是通过简单的步骤将数据处理为可操作的直观内容，本书作者还会告诉你这是如何完成的，甚至比你想象的容易。通过一个简单的教程，你将学习如何通过 R 语言构建可以解决商业问题的深度神经网络模型。一旦你掌握了该教程，就可以很容易地将自己的知识转化为符合实际需求的商业应用。

事半功倍：R 语言简单易用，并且可以在所有主流操作系统上免费使用！本书的每一章都按部就班地介绍了深度神经网络的不同方面。在解决一些具有挑战性的实际业务问题时，你需要亲自动手实践一下。

业务分析迅速

本书是一本介绍深度学习在商业分析方面的入门读物，它将以通俗易懂的文字为你介绍如何利用深度学习提升商业价值。

你将学到以下知识：

- 挖掘深度神经网络于在线新闻报道分类方面的潜力；
- 开发用于评估客户流失的解决方案；
- 为客户品牌选择建模设计一些成功的应用程序；
- 掌握有效的产品需求预测技术；
- 部署深度神经网络以预测信用卡消费情况；
- 采用成功的解决方案预测汽车价格。

无痛学习：本书是专门为那些希望在短时间内掌握如何使用深度神经网络

络的人而量身打造的。它利用强大的 R 编程语言为读者提供必要的工具，从而最大限度地提高读者的认知，加深读者对上述知识的理解，继而提升其在数据分析项目中的应用能力。

快马加鞭：如果你希望加快学习进度并学以致用，那么本书就是一个很好的入门方案，它揭示了深度神经网络的工作原理，并为你提供了一个易于遵循的实践指南，告诉你如何使用强大且免费的 R 编程语言快速地构建深度神经网络。

轻松入门：本书包含了新手入门所需的一切。它是属于你的实践手册和战术指南，只需按部就班地遵循其中的步骤就有可能从新手迈入专家的行列。

学习本书，意味着你使用深度神经网络方面的技能将获得重大突破。

N. D. Lewis 的其他著作

- Deep Learning Step by Step with Rython
- Deep Learning Made Easy with R
 - Volume I: A Gentle Introduction for Data Science
 - Volume II: Practical Tools for Data Science
 - Volume III: Breakthrough Techniques to Transform Performance
- Build Your Own Neural Network TODAY
- 92 Applied Predictive Modeling Techniques in R
- 100 Statistical Tests in R
- Visualizing Complex Data Using R
- Learning from Data Made Easy with R

前言

现在市面上深度学习方面的图书越来越多，那么为什么还有人要继续写另外的呢？我一直对深度学习的概念及其商业应用很着迷。它提供了很多新的可能性，可以从数据中提取很多有用的见解。现在，随着 R 语言的兴起，这些工具比以往更容易使用。很多知名的科技企业已经部署了深度学习技术来帮助它们解决业务问题。但是，现在很多教科书侧重于技术细节，并且篇幅过长，偏数学化，对于应用案例的讲述反而太少。

本书的目标是向读者介绍深度神经网络中各种强大的应用工具。我的愿望是帮助读者通过 R 语言开发得心应手的实用工具。本书通过使用真实数据和 R 编程语言阐述了深度学习技术在解决商业问题上的潜力，它采用通俗易懂的语言避免涉及过多的数学知识。

你不需要是一名天才

本书内容通俗易懂，读者不需要是一名专业的统计员或编程专家，就能理解本书讨论的实用理念并能够直接从解决方案中受益。对于希望将深

度学习工具应用于商业问题的读者来说，目前市面上有关这方面的书籍仍难觅踪迹。本书的主要目标是鼓励读者广泛采用深度学习技术来解决企业面临的问题。

内容概要

本书包含了构建深度神经网络的所有重要内容，从具体的商业插图，到按部就班的 R 示例，再到读者可能在实际工作中遇到的问题。本书收集了经过精简处理的最佳实用工具、理念和可能用到的提示技巧，其中包括一系列真实的商业分析案例，它们用于说明业务问题的特殊性和深度神经网络技术用于解决这类问题的强大特性。

实践方法是贯穿本书的重点，它们已被证明能够提供出色的性能，并能够提高清晰度和效率。出于这个原因，本书在相关章节中包含了很多技术细节。

关于 R 新手

本书旨在为读者提供一条捷径，它将逐步向读者展示如何在免费且流行的 R 语言统计软件包中构建各种模型。

这些示例都有明确的说明，并且可以直接在 R 语言开发环境下输入并显示结果。R 语言新手也可以轻松地使用本书，而无需任何预备知识。最好通过输入示例中的代码并阅读代码中的注释来理解它们。可以在 R Project 网站免费下载适用于初学者的 R 语言程序和教程。如果读者对 R 语言一无所知，可前往 CRAN 网站查看 R 语言的简易教程，这些教程对于 R 语言新手真的非常棒。

最后，数据科学最终是关乎现实生活和真实的人的，也是力求将机器学习算法应用于实际问题，以便提供有用的解决方案。无论你是谁，来自何处，你的背景或学历是什么，你都有能力掌握本书介绍的方法。有了适当的软件工具、足够的耐心和正确的指导，我个人认为深度学习方法可以被任何对它感兴趣的人所掌握。

古希腊哲学家伊壁鸠鲁曾经说过：

“我写东西并不是为了取悦大众，而是为了你，我们每个人都有足够的观众。”

尽管本书中的想法可能会传递给万千读者，但是我仍然努力遵循伊壁鸠鲁的原则——让你阅读的每一页内容都非常有意义。

一个建议

当读完本书的时候，希望你将能够把本书讨论过的某些方法付诸实践，你会因为在 R 中能够快速简单地将这些技术应用和部署而感到惊讶。只需几次实战应用，你将很快成为一名熟练的执业人员。因此，建议你将这些阅读到的内容付诸实践。

Dr. N.D. Lewis

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

要获得本书的配套资源，请在异步社区本书页面中单击 **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。

如果您是教师，希望获得教学配套资源，请在社区本书页面中直接联系本书的责任编辑。

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，单击“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](#)

页码: 页内位置(行数): 勘误次数:

B I U

字数统计

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 如何充分利用本书	1	异彩	20
1.1 软件包使用建议	2	3.2 还有谁希望快速地了解 深度学习的强大	21
1.2 高效使用函数	3	3.3 改进价值创造链	23
1.3 无需等待	3	3.4 如何进行智能化营销	24
1.3.1 勤于动手	4	3.5 客户流失——以及如何 增加利润的小技巧	25
1.3.2 深度学习的价值	4	3.6 挖掘预测产品需求过程中 隐藏的商机	25
1.4 参考资料	5	3.7 参考资料	27
第 2 章 商业分析与神经网络	6	第 4 章 神经元和激活函数	29
2.1 数据价值创造周期	8	4.1 人工神经元简介	30
2.2 神经网络简介	9	4.2 激活函数	31
2.3 模式识别的本质	10	4.3 简化数学计算	32
2.3.1 时序模式	12	4.4 S 型激活函数简介	33
2.3.2 复杂的商业模式	14	4.5 参考资料	34
2.4 属性、分类和回归	15	第 5 章 神经网络的学习机制	36
2.4.1 属性	15	5.1 反向传播算法简介	37
2.4.2 回归	16	5.2 基本算法的工作原理	38
2.4.3 分类	16	5.3 关于渐变下降	39
2.5 参考资料	17	5.4 误差面简介	40
第 3 章 商业中的深度学习	19		
3.1 古典游戏让深度学习大放			

5.4.1	均方根误差	40			
5.4.2	局部极小值	40			
5.5	关于随机梯度下降的 注意事项	43			
5.6	参考资料	43			
第 6 章	深度神经网络简介	44			
6.1	关于精确预测的常识	45			
6.2	在没有知识储备的情况 下为复杂的数学关系 建模	51			
6.3	整合深度神经网络 蓝图	53			
6.4	深度神经网络的完美 剖析	53			
6.5	选择最佳层数	54			
6.6	参考资料	55			
第 7 章	在线热点新闻分类	57			
7.1	在线新闻的特点	58			
7.2	如何从网上下载在线 新闻样本	59			
7.3	一种浏览数据样本的 简单方法	61			
7.4	如何预处理新闻转发的 频率	63			
7.5	标准化的重要性	64			
7.6	创建训练样本	66			
7.7	适合深度神经网络的 证明方法	67			
7.8	分类预测	69			
7.9	需要包含多少个神经元 的答案	70			
7.9.1	一个关键点	71			
7.9.2	核心思想	72			
7.10	构建一个更复杂的 模型	72			
7.11	混淆矩阵	73			
7.11.1	table 函数	74			
7.11.2	测试集上的 性能	75			
7.12	实践出真知	76			
7.13	参考资料	76			
第 8 章	为客户流失建模以促进 业务增长	78			
8.1	客户流失的原因	79			
8.2	电信行业的客户流失	81			
8.3	如何将客户流失样本下载 到本地硬盘	83			
8.4	一种收集数据和查看特征 的简单方法	83			
8.4.1	转换因子变量	84			
8.4.2	转换响应变量	85			
8.4.3	清理	86			
8.4.4	查看数据	86			
8.5	快速构建一个深度神经 网络	87			
8.6	接收器操作特性曲线下的 面积	88			
8.7	Tanh 激活函数	91			
8.8	关于学习率	93			
8.9	动量的完整直观指南	95			

8.9.1 选择动量值	96	9.8 重新训练一个替代模型	124
8.9.2 R 中的学习率和动量	97	9.9 如何选择重复的次数	125
8.10 不平衡类的问题	98	9.9.1 一个问题和一个答案	126
8.11 一种易用的不平衡类解决方案	100	9.9.2 查看多次重复的性能	127
8.11.1 unbalanced 软件包	101	9.10 一个建模错误可以严重影响性能	128
8.11.2 运行模型	102	9.11 简单模型如何提供稳定的性能	130
8.11.3 选择模型	103	9.12 实践出真知	131
8.11.4 每种模型的测试集性能	104	9.13 参考资料	132
8.12 实践出真知	106	第 10 章 预测客户信用卡消费的艺术	133
8.13 参考资料	106	10.1 明确信贷的角色	134
第 9 章 产品需求预测	108	10.2 信用卡数据	135
9.1 自行车共享系统	109	10.3 预处理样本数据	137
9.2 数据样本的分布和相关性	110	10.3.1 处理因子属性	138
9.2.1 目标变量	113	10.3.2 处理数值变量	139
9.2.2 预备属性	114	10.3.3 创建核心数据帧	139
9.2.3 属性规范化	116	10.4 一种设计深度神经网络的简单方法	140
9.2.4 训练集	117	10.4.1 一种替代模型	141
9.3 自动化公式生成	117	10.4.2 一种解释	142
9.4 弹性反向传播解密	119	10.5 过度训练的挑战	144
9.5 奥卡姆剃刀法则的解释	119	10.6 提早停止的简单策略	146
9.6 如何使用奥卡姆剃刀法则	120	10.7 实践出真知	147
9.7 确定性能基准的简单方法	123		

10.8	参考资料	148	12.1	二手车价格的关键因素	172
第 11 章	客户品牌选择建模简介	151	12.2	下载二手车数据集	173
11.1	品牌选择的概念性框架	152	12.3	评估二手车价格和其他属性的关系	174
11.1.1	效用的挑战	152	12.4	一个简单的数据预处理技巧	178
11.1.2	实际情况	153	12.5	快速建立训练集和测试集	180
11.2	检查样本数据	154	12.6	充分利用 mini batching 算法	180
11.3	S 型激活函数的实际限制	156	12.6.1	一个问题	181
11.3.1	模型精度	157	12.6.2	R 中的批量大小	182
11.3.2	消失的梯度	158	12.7	测量和评估模型性能	182
11.4	深度学习工具箱中必备的一个激活函数	159	12.8	高效交叉验证的基本要领	183
11.5	保持技术的秘密	160	12.9	一个可以轻松模拟的实用示例	185
11.5.1	一个实际问题	161	12.9.1	一个简单的 for 循环	185
11.5.2	重要提示	163	12.9.2	测试集上的性能	187
11.6	数据预处理的魔力	164	12.10	最后的思考	188
11.6.1	继续研究消费者品牌选择模型	164	12.11	参考资料	189
11.6.2	测试集性能	166			
11.6.3	修改测试——训练集分割	167			
11.7	实践出真知	168			
11.8	参考资料	168			
第 12 章	汽车价格预测	171			