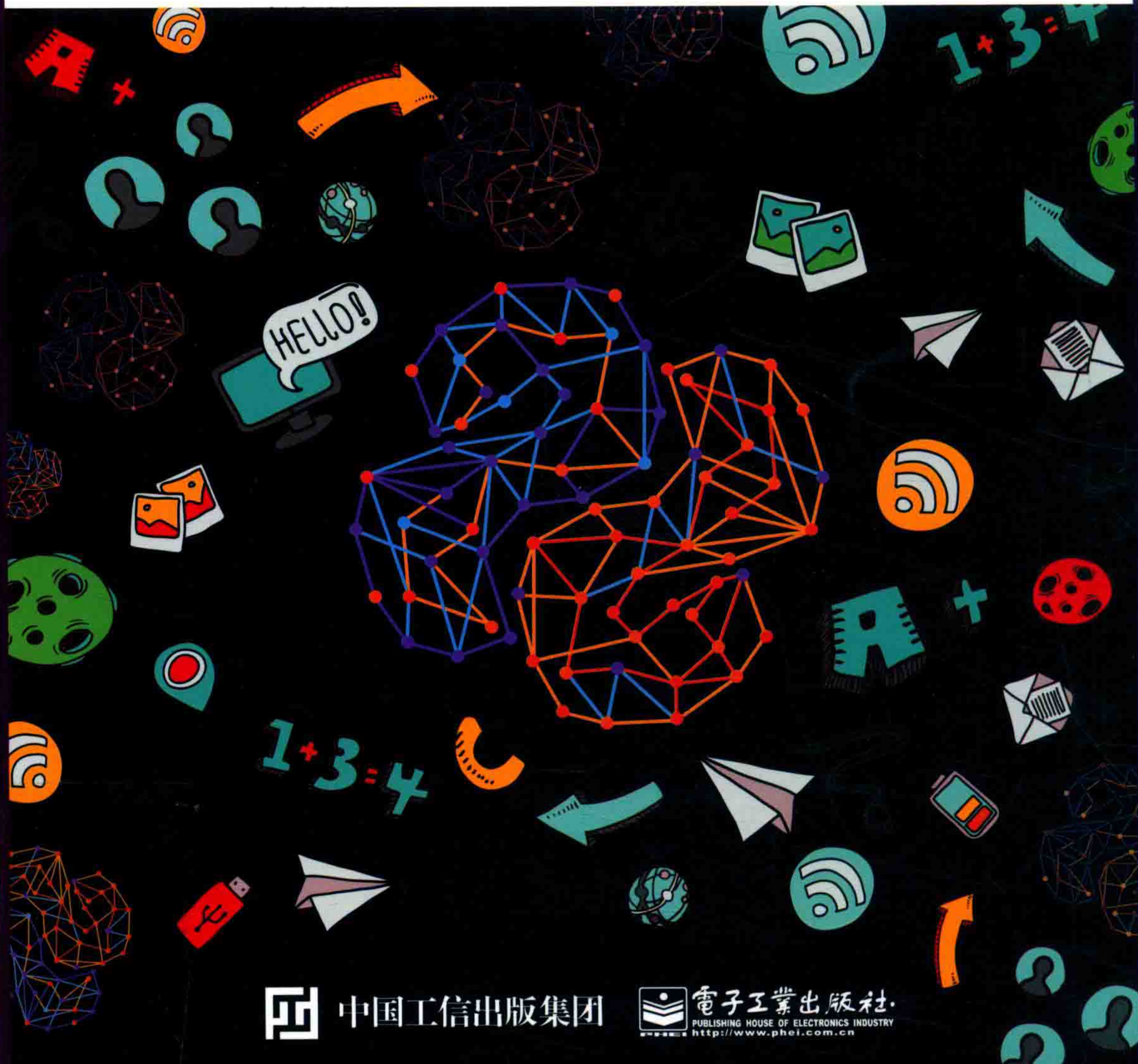


Broadview[®]
www.broadview.com.cn

深入浅出深度学习

原理剖析与Python实践

黄安埠 著



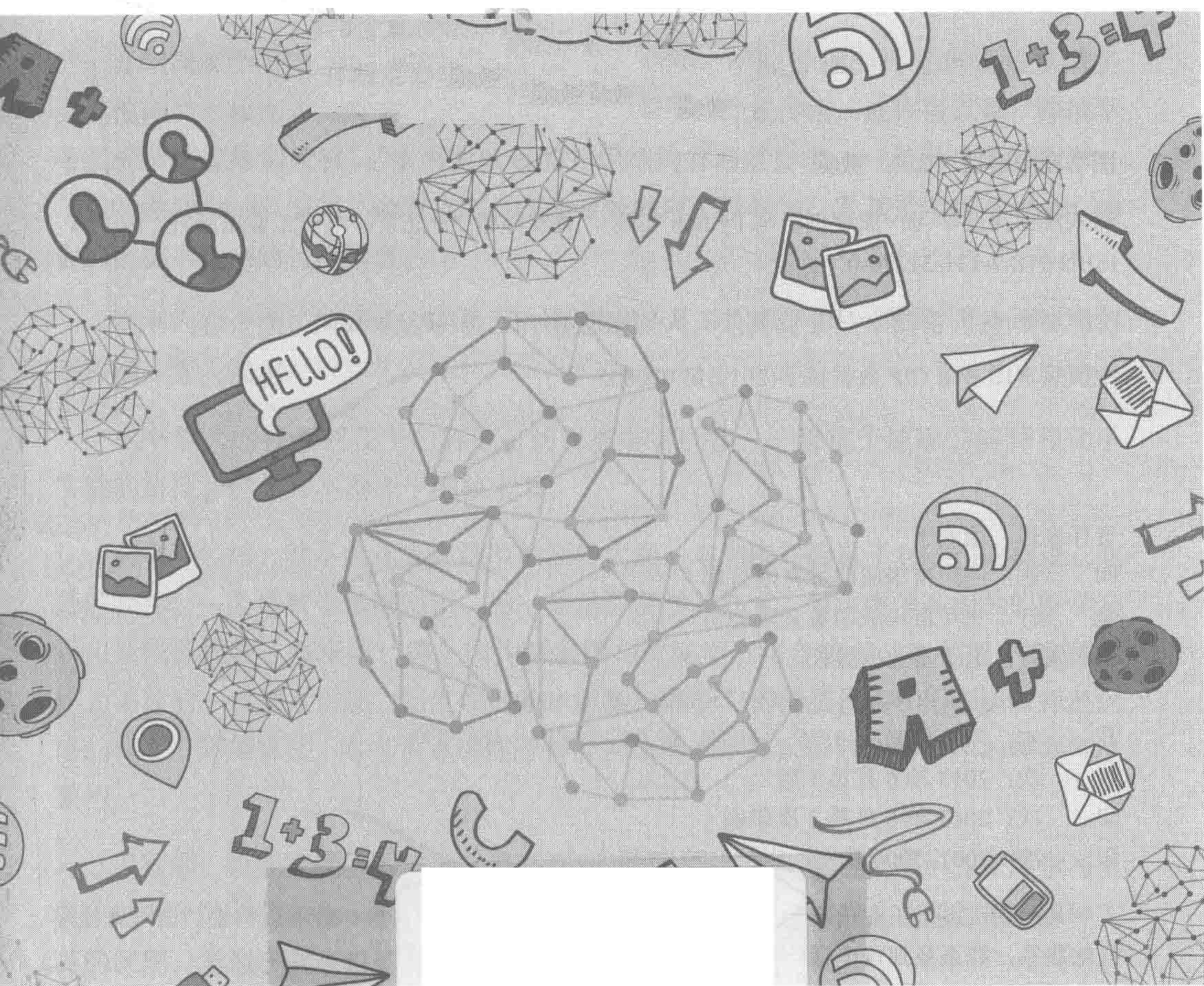
 中国工信出版集团

 电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

深入浅出深度学习

原理剖析与Python实践

黄安埠 著



电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书介绍了深度学习相关的原理与应用，全书共分为三大部分，第一部分主要回顾了深度学习的发展历史，以及 Theano 的使用；第二部分详细讲解了与深度学习相关的基础知识，包括线性代数、概率论、概率图模型、机器学习和最优化算法；在第三部分中，针对若干核心的深度学习模型，如自编码器、受限玻尔兹曼机、递归神经网络和卷积神经网络等进行详细的原理分析与讲解，并针对不同的模型给出相应的具体应用。

本书适合有一定高等数学、机器学习和 Python 编程基础的在校学生、高校研究者或在企业中从事深度学习的工程师使用，书中对模型的原理与难点进行了深入分析，在每一章的最后都提供了详细的参考文献，读者可以对相关的细节进行更深入的研究。最后，理论与实践相结合，本书针对常用的模型分别给出了相应的应用，读者也可以在 Github 中下载和查看本书的代码 (<https://github.com/innovation-cat/DeepLearningBook>)。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目 (CIP) 数据

深入浅出深度学习：原理剖析与 Python 实践 / 黄安埠著. —北京：电子工业出版社，2017.6
ISBN 978-7-121-31270-0

I. ①深… II. ①黄… III. ①软件工具—程序设计 IV. ①TP311.561

中国版本图书馆 CIP 数据核字(2017)第 070055 号

责任编辑：徐津平

印 刷：三河市华成印务有限公司

装 订：三河市华成印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：720×1000 1/16 印张：22.25 字数：401 千字

版 次：2017 年 6 月第 1 版

印 次：2017 年 6 月第 2 次印刷

印 次：3001~7000 册 定价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888, 88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819, faq@phei.com.cn。

推荐序 1

介绍深度学习的书籍不少，但是《深入浅出深度学习：原理剖析与 Python 实践》与其他同类书相比，视角明显不同。如果要给本书写个宣传语，或许可以是“深度学习工程师速成培训教材”。本书对读者的知识结构有两点要求：一是学过高等数学，二是熟悉 Python 编程。换言之，各个专业的理工科学生，尤其是学过 Python 编程的，都是此书的目标读者。

本书内容全面，但是取舍明确，有重点地深入，尤其对于技术的重点难点解释得很详细，深入浅出。

本书最大的特色就在于内容取舍的尺度非常明确——着重于原理解释和动手实践的路径，但是并不拘泥于细枝末节。

胸中有经纬，就不会迷失在细节的汪洋大海。本书刚好侧重于对经纬的梳理。此处的经纬，一类是数学基础知识，另一类是深度学习技术。与深度学习相关的数学知识包括线性代数、概率统计等。因为概率图与深度学习结合较多，本书把“概率图”作为单独的一章重点讲述。深度学习技术包括机器学习的传统技术、用于训练神经网络的梯度下降等算法。本书重点讲述了神经网络的基本算法以及几种常用的深度网络架构。

工程师，重在实践。工欲善其事必先利其器，实践深度学习，离不开深度学习工具。本书介绍了 Theano 工具集的基本用法。其实深度学习工具，一通百通，各种工具的区别，类似于北京口音与东北口音的区别。

认真读完此书，读者应该拥有三项能力：一是读得懂深度学习的论文；二是读得

懂深度学习的代码；三是能够自行开发简单的深度学习应用。

总结一下，在深度学习技术异常火爆，深度学习工程师奇缺的当下，如何快速培养深度学习方向的工程师，是一个迫切的问题。此书是难得的好教材。

邓侃博士

邓侃，美国卡内基梅隆大学(CMU)计算机学院博士，专攻人工智能和数据挖掘。历任美国甲骨文(Oracle)主任系统架构师，美国泰为手机导航公司(Telenav)北京分公司总经理，百度网页搜索高级总监。2015年至今创建北京大数医达科技有限公司，专注于研发医疗大数据和人工智能医生，出任CTO。

推荐序 2

在过去的这十年，深度学习已经席卷了整个科技界和工业界，2016 年谷歌阿尔法狗打败围棋世界冠军李世石，更是使其成为备受瞩目的技术焦点。记得 2010 年 7 月我参与了 Facebook 人脸识别（Face Detection）工作，那时候深度学习还没有普及，这个系统是当时世界上最大的实时人脸识别系统，每天都有几千万张脸被找出来。随着硬件的成熟和数据的指数级增加，深度学习在很多问题上成为人工智能最火和最有效的方法。

这几年，我身边越来越多的计算机从业者，甚至在校大学生向我咨询，有效了解和学习这个当前人工智能最热门的领域的方法。没错，一方面深度学习很热门，人人都对此充满了好奇和向往，但是同时，它对于初学者往往挑战也不小，因为需要一定的数学基础，同时各种专业术语，如果不用深入浅出的方式讲，很容易让人产生畏难情绪。

所以我推荐黄安埠的这本《深入浅出深度学习：原理剖析与 Python 实践》给所有对深度学习有兴趣的朋友们，本书涵盖了深度学习的理论以及各种常见的深度学习模型，你们会在这本书中找到关于深度学习最实用的知识和信息。对于初学者来说，这是一本非常通俗易懂的入门教材。

同时我也推荐该书给深度学习领域的开发者和数据科学家，因为作者也分享和总结了许多深度学习的最佳实践法，并辅以相当多的实际应用案例加以诠释，是一本值得借鉴参考的好书。

陈尔东

前 Twitter 总监，前 Facebook 经理和早期工程师

前言

What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.

—— Marvin Minsky

智能（Intelligence）这个词的出现最早可以追溯到古希腊时期，当时人们已经开始梦想能创造出一种像人类一样，具有独立思考和推理能力的机器，但由于受到当时生产力水平的制约，古人对“智能”的研究更多的是停留在理论探索阶段。到了近代，尤其是具有划时代意义的达特茅斯会议的召开，标志着人工智能开始从理论探索进入到理论与应用相结合的实践阶段。从20世纪50年代开始，人工智能的发展大致经历了三个阶段，分别从最初的逻辑推理，到统计机器学习，再到近年来逐渐占据主流地位的深度学习。

虽然深度学习是一门以神经网络为核心的学科，但人们普遍认为深度学习始于2006年，当时Hinton等人提出基于深度置信网络（DBN）逐层预训练的方法来训练深层模型，并首次提出了深度学习的概念。此后，深度学习开始进入人们的视野，但那时候深度学习更多的是少数顶尖科学家研究的领域，并没有得到大规模的应用和推广。直到2012年，Hinton和他的两个学生Alex Krizhevsky、Illya Sutskever，将卷积神经网络应用到ImageNet竞赛中，并取得了分类错误率15%的成绩，这个成绩比第二名低了近11个百分点，这一历史性的突破，使得人们开始意识到深度学习所拥有的巨大潜力，在这之后，深度学习开始在工业界，尤其是计算机视觉、语音识别和自然语言处理等领域，大规模应用，并且取得了比以往更好的效果。到了2016年，随着AlphaGo的横空出世，它的惊人表现将深度学习的热度推向了顶峰，因此2016年也被很多学者认

为是人工智能元年，事实上，当前人工智能已经影响到人们生活的各个方面，如语音助手、语音搜索、无人驾驶汽车、人脸识别等，为人们的生活带来了极大的方便，人工智能也必将在今后相当长的一段时间内，继续推动着人类的技术发展。

在本书编写的过程中，市面上有关深度学习方面的中文书籍较少，因此作者希望能从理论和应用相结合的角度，对深度学习的相关知识进行较为全面的梳理，本书既可以作为初级读者的入门书籍，也适合中级读者用来加深对理论知识的理解。本书覆盖了线性代数、概率论、数值计算与最优化等基础知识，以及深度学习的两大核心：概率图模型和深度神经网络。具体来说，本书由以下三大部分构成：

第1部分是概要，共分为两章。第1章主要阐述了深度学习、人工智能相关的背景，深度学习的原理，以及当前流行的深度学习框架对比；第2章介绍了深度学习框架Theano的使用，着重对Theano的基础知识和编程范式进行了讲解。

第2部分是与深度学习相关的数学和机器学习方面的基础知识，共分为5章。第3章介绍线性代数基础知识；第4章介绍了概率论和数理统计相关的知识；第5章介绍概率图模型，包括贝叶斯网络和马尔科夫网络的原理；第6章简要回顾机器学习的基础知识，并介绍机器学习模型与深度学习模型之间的联系；第7章，深入分析几种常用的机器学习最优化方法，包括具有一阶收敛速度的梯度下降法和共轭梯度法，以及具有二阶收敛速度的牛顿法和拟牛顿法。

第3部分介绍了各种常见的深度学习模型，包括一系列的深度学习模型理论及其应用，本部分共分为6章。第8章介绍全连接前馈神经网络，包括网络结构和激活函数的相关知识；第9章将深入分析反向传播算法，以及梯度消失问题。梯度消失也是深度神经网络训练的一大难点，我们将介绍当前有效解决深度网络训练中过拟合和欠拟合的常见技巧，包括Batch Normalization、残差网络、Dropout等；第10章介绍本书的第一种无监督网络模型：自编码器及其变种模型；第11章介绍一种深度概率图模型——受限玻尔兹曼机，与自编码器一样，受限玻尔兹曼机也是一种常见的无监督网络模型，最后介绍如何将受限玻尔兹曼机应用于个性化推荐领域中；第12章，将介绍一种应用非常广泛的网络结构——递归神经网络，深入分析递归网络的结构及其变形网络，如LSTM、GRU等，并以语言模型为例，介绍递归神经网络在自然语言处理中的应用；第13章介绍另一种常见的模型结构：卷积神经网络，包括卷积网络的卷积层和池化层结构设计，以及其在文本分类中的应用。

关于本书的源代码，读者也可以从Github上（<https://github.com/innovation->

cat/DeepLearningBook) 下载查看。深度学习近年来处于高速发展的阶段，很多更先进的理论和算法正被不断提出，因此本书无法覆盖所有的模型与算法，加之作者水平和精力所限，书中难免有错漏之处，承蒙各位读者不吝告知，如对本书有任何疑问或建议，读者可以通过邮箱 huanganbu@gmail.com 给我反馈。

在本书的撰写过程中，得到了很多行业专家和好友的支持，在此，特别感谢香港科技大学计算机系主任杨强教授、原百度网页搜索高级总监邓侃博士、原Twitter工程总监陈尔东先生，感谢他们在百忙之中抽时间审阅我的书稿，提出了很多宝贵的意见，并为我写下推荐序。

在本书的撰写过程中，还得到了电子工业出版社刘皎编辑和汪达文编辑的极大帮助，在此表示衷心的感谢；感谢我在腾讯公司的上级李深远先生对我工作的支持，也感谢其他各位关心我工作的朋友和同事。

最后，非常感谢我的家人对我工作的理解和支持，他们在我写作的过程中给予了很大的照顾和鼓励，也是促使我能完成本书写作的最大动力。

黄安埠

2017年3月于深圳

轻松注册成为博文视点社区用户 (www.broadview.com.cn)，扫码直达本书页面。

- **下载资源**：本书如提供示例代码及资源文件，均可在 [下载资源](#) 处下载。
- **提交勘误**：您对书中内容的修改意见可在 [提交勘误](#) 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 [读者评论](#) 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/31270>



目录

第 1 部分 概要	1
1 绪论.....	2
1.1 人工智能、机器学习与深度学习的关系	3
1.1.1 人工智能——机器推理	4
1.1.2 机器学习——数据驱动的科学	5
1.1.3 深度学习——大脑的仿真	8
1.2 深度学习的发展历程	8
1.3 深度学习技术概述	10
1.3.1 从低层到高层的特征抽象	11
1.3.2 让网络变得更深	13
1.3.3 自动特征提取	14
1.4 深度学习框架	15
2 Theano 基础.....	19
2.1 符号变量.....	20
2.2 符号计算的抽象——符号计算图模型	23
2.3 函数	26
2.3.1 函数的定义	26
2.3.2 Logistic回归	27
2.3.3 函数的复制	29
2.4 条件表达式.....	31

2.5	循环.....	32
2.6	共享变量.....	39
2.7	配置.....	39
2.7.1	通过THEANO_FLAGS配置.....	40
2.7.2	通过.theanorc文件配置.....	41
2.8	常用的Debug技巧.....	42
2.9	小结.....	43
第2部分 数学与机器学习基础篇.....		45
3	线性代数基础.....	46
3.1	标量、向量、矩阵和张量.....	46
3.2	矩阵初等变换.....	47
3.3	线性相关与向量空间.....	48
3.4	范数.....	49
3.4.1	向量范数.....	49
3.4.2	矩阵范数.....	53
3.5	特殊的矩阵与向量.....	56
3.6	特征值分解.....	57
3.7	奇异值分解.....	58
3.8	迹运算.....	60
3.9	样例：主成分分析.....	61
4	概率统计基础.....	64
4.1	样本空间与随机变量.....	65
4.2	概率分布与分布函数.....	65
4.3	一维随机变量.....	66
4.3.1	离散随机变量和分布律.....	66
4.3.2	连续随机变量和概率密度函数.....	67
4.4	多维随机变量.....	68
4.4.1	离散型二维随机变量和联合分布律.....	69
4.4.2	连续型二维随机变量和联合密度函数.....	69
4.5	边缘分布.....	70
4.6	条件分布与链式法则.....	71

4.6.1	条件概率.....	71
4.6.2	链式法则.....	73
4.7	多维随机变量的独立性分析.....	73
4.7.1	边缘独立.....	74
4.7.2	条件独立.....	74
4.8	数学期望、方差、协方差.....	75
4.8.1	数学期望.....	75
4.8.2	方差.....	76
4.8.3	协方差.....	76
4.8.4	协方差矩阵.....	78
4.9	信息论基础.....	81
4.9.1	信息熵.....	81
4.9.2	条件熵.....	83
4.9.3	互信息.....	84
4.9.4	相对熵与交叉熵.....	84
5	概率图模型.....	87
5.1	生成模型与判别模型.....	89
5.2	图论基础.....	90
5.2.1	图的结构.....	90
5.2.2	子图.....	91
5.2.3	路径、迹、环与拓扑排序.....	92
5.3	贝叶斯网络.....	95
5.3.1	因子分解.....	96
5.3.2	局部马尔科夫独立性断言.....	99
5.3.3	I-Map与因子分解.....	100
5.3.4	有效迹.....	103
5.3.5	D-分离与全局马尔科夫独立性.....	108
5.4	马尔科夫网络.....	108
5.4.1	势函数因子与参数化表示.....	109
5.4.2	马尔科夫独立性.....	111
5.5	变量消除.....	114
5.6	信念传播.....	116
5.6.1	聚类图.....	116

5.6.2	团树.....	120
5.6.3	由变量消除构建团树.....	123
5.7	MCMC采样原理.....	126
5.7.1	随机采样.....	127
5.7.2	随机过程与马尔科夫链.....	128
5.7.3	MCMC采样.....	132
5.7.4	Gibbs采样.....	134
5.8	参数学习.....	137
5.8.1	最大似然估计.....	137
5.8.2	期望最大化算法.....	138
5.9	小结.....	140
6	机器学习基础.....	142
6.1	线性模型.....	143
6.1.1	线性回归.....	143
6.1.2	Logistic回归.....	148
6.1.3	广义的线性模型.....	150
6.2	支持向量机.....	151
6.2.1	最优间隔分类器.....	152
6.2.2	对偶问题.....	155
6.2.3	核函数.....	156
6.3	朴素贝叶斯.....	160
6.4	树模型.....	162
6.4.1	特征选择.....	163
6.4.2	剪枝策略.....	165
6.5	聚类.....	166
6.5.1	距离度量.....	167
6.5.2	层次聚类.....	168
6.5.3	K-means聚类.....	171
6.5.4	谱聚类.....	172
7	数值计算与最优化.....	177
7.1	无约束极小值的最优化条件.....	177
7.2	梯度下降.....	179

7.2.1	传统更新策略.....	181
7.2.2	动量更新策略.....	183
7.2.3	改进的动量更新策略.....	184
7.2.4	自适应梯度策略.....	187
7.3	共轭梯度.....	188
7.4	牛顿法.....	192
7.5	拟牛顿法.....	194
7.5.1	拟牛顿条件.....	194
7.5.2	DFP算法.....	195
7.5.3	BFGS算法.....	196
7.5.4	L-BFGS算法.....	197
7.6	约束最优化条件.....	200
第3部分 理论与应用篇.....		205
8	前馈神经网络.....	206
8.1	生物神经元结构.....	207
8.2	人工神经元结构.....	208
8.3	单层感知机.....	209
8.4	多层感知机.....	212
8.5	激活函数.....	217
8.5.1	激活函数的作用.....	217
8.5.2	常用的激活函数.....	219
9	反向传播与梯度消失.....	225
9.1	经验风险最小化.....	227
9.2	梯度计算.....	228
9.2.1	输出层梯度.....	228
9.2.2	隐藏层梯度.....	230
9.2.3	参数梯度.....	234
9.3	反向传播.....	235
9.4	深度学习训练的难点.....	237
9.4.1	欠拟合——梯度消失.....	237
9.4.2	过拟合.....	240

10	自编码器及其相关模型	243
10.1	自编码器.....	243
10.2	降噪自编码器.....	245
10.3	栈式自编码器.....	247
10.4	稀疏编码器.....	250
10.5	应用：cifar10图像分类	254
11	玻尔兹曼机及其相关模型	258
11.1	玻尔兹曼机.....	258
11.2	能量模型.....	261
11.2.1	能量函数.....	261
11.2.2	从能量函数到势函数.....	262
11.2.3	从势函数到概率分布.....	263
11.3	推断	264
11.3.1	边缘分布.....	265
11.3.2	条件分布.....	267
11.4	学习	270
11.4.1	最大似然估计	271
11.4.2	对比散度.....	274
11.5	应用：个性化推荐	276
11.5.1	个性化推荐概述.....	276
11.5.2	个性化推荐架构与算法.....	279
11.5.3	RBM与协同过滤.....	285
12	递归神经网络.....	291
12.1	Elman递归神经网络.....	292
12.2	时间反向传播.....	295
12.3	长短时记忆网络	299
12.4	结构递归神经网络	302
12.5	应用：语言模型	308
12.5.1	N元统计模型	308
12.5.2	基于 LSTM 构建语言模型	312

13	卷积神经网络	318
13.1	卷积运算.....	319
13.2	网络结构.....	320
13.3	卷积层.....	324
13.4	池化层.....	329
13.5	应用：文本分类.....	333

第 1 部分

概 要
