

关于深度学习的导论性著作，也是了解深度学习的入门之书。

详述了深度学习的9大重要模型及其学习算法、变种模型和混杂模型，内容翔实，具有提纲挈领的指导意义。

基于Matlab、Python和C++相关的程序案例介绍深度学习模型，有助于读者全面了解深度学习模型和算法的实现途径。



深度学习导论 及案例分析

李玉鑑 张婷 等著



机械工业出版社
China Machine Press



深度学习导论 及案例分析

李玉鑑 张婷 等著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

深度学习导论及案例分析 / 李玉鑑等著. —北京: 机械工业出版社, 2016.10

ISBN 978-7-111-55075-4

I. 深… II. 李… III. 学习系统—研究 IV. TP273

中国版本图书馆 CIP 数据核字 (2016) 第 236304 号

深度学习是近年来在神经网络发展史上掀起的一波新浪潮, 是机器学习的一大热点方向, 因在手写字符识别、维数约简、图像理解和语音处理等方面取得了巨大进展, 所以很快受到了学术界和工业界的高度关注。在本质上, 深度学习就是对具有深层结构的网络进行有效学习的各种方法。本书介绍了深度学习的起源和发展, 强调了深层网络的特点和优势, 说明了判别模型和生成模型的相关概念, 详述了深度学习的 9 种重要模型及其学习算法、变种模型和混杂模型, 讨论了深度学习在图像处理、语音处理和自然语言处理等领域的广泛应用, 总结了深度学习目前存在的问题、挑战和未来的发展趋势, 还分析了一系列深度学习的基本案例。本书可以作为计算机、自动化、信号处理、机电工程、应用数学等相关专业的研究生、教师和科研工作者在具备神经网络基础知识后, 进一步了解深度学习理论和方法的入门教材或导论性参考书, 有助于读者掌握深度学习的主要内容并开展相关研究。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 张梦玲

责任校对: 董纪丽

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2016 年 10 月第 1 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 18.75

书 号: ISBN 978-7-111-55075-4

定 价: 59.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光/邹晓东

P R E F A C E

前 言

“深度学习”一词大家已经不陌生了，随着在不同领域取得了超越其他方法的成功，深度学习在学术界和工业界掀起了一次神经网络发展史上的新浪潮。运用深度学习解决实际问题，不仅是学术界高素质人才所需的技能，而且是工业界商业巨头进行竞争的核心武器。为适应这一发展的需要，作者以长期的相关研究和教学工作为基础，经过2~3年的调研和努力，终于编写完本书。这是一本关于深度学习的入门教材和导论性参考书，受众对象包括计算机、自动化、信号处理、机电工程、应用数学等相关专业的研究生、教师和科研工作者，本书有助于他们在具备神经网络的基础知识后进一步了解深度学习的理论和方法。

自2006年诞生以来，深度学习很快成长壮大，并有一些相关的英文书籍陆续出版。虽然国内也开始出现译著，但对深度学习的内容概括得并不全面，远不能够满足市场需求。本书的内容几乎涵盖了深度学习的所有重要方面，结构上分为基础理论和案例分析两个部分。在基础理论部分，本书不仅介绍了深度学习的起源和发展、特点和优势，而且描述了深度学习的9种重要模型，包括受限玻耳兹曼机、自编码器、深层信念网络、深层玻耳兹曼机、和积网络、卷积神经网络、深层堆叠网络、循环神经网络和长短时记忆网络。此外，还讨论了这些模型的学习算法、变种模型和混合模型，以及它们在图像视频处理、音频处理和自然语言处理等领域中的广泛应用，并总结了有关的开发工具、问题和挑战。在案例分析部分，本书主要挑选了一些深度学习的程序案例进行细致的说明和分析，指导读者学习有关的程序代码和开发工具，以便在解决实际问题时加以灵活利用。其中，每个程序案例都包括模块简介、运行过程、代码分析和使用技巧这4个部分，层次结构清晰，以利于读者选择和学习，并在应用中拓展思路。本书的一个不足之处是：案例分析部分没有涉及“和积网络”和“深层堆叠网络”，这是因为和积网络的运行需要大规模集群的硬件条件，另外也很难找到便于构造深层堆叠网络案例的程序代码。

本书的一大特色是从初学者的角度出发，在知识结构的布局上注重深入浅出，对深度学习

的模型涵盖得较全面，文献引用非常丰富，既适合读者入门学习，又有助于他们深入钻研。同时，本书也试图纠正许多读者对深度学习的一些错误理解，比如认为多层感知器不是深度学习模型，认为自编码器能够直接用来识别手写字符，认为受限玻耳兹曼机也是严格意义上的深度学习模型，等等。

本书的另一个特色是通过程序案例介绍深度学习模型。这对缺乏相关背景知识的读者可能非常有帮助，使他们在知其然不知其所以然的情况下运行深度学习程序并获得计算结果，从而在积累实践经验和感性认识的过程中逐步了解深度学习的有关内容。本书的案例涉及三种常见的编程语言：Matlab、Python 和 C++。其中，很多深度学习程序是用 Matlab 编写的，可以直接运行。如果使用 Python 语言编写深度学习程序，则可以调用 Theano 开源库；若使用 C++ 语言，则可以调用 Caffe 开源库。不同的语言分析案例有助于读者全面了解深度学习模型和算法的实现途径，并根据自己的熟练程度灵活选择。

本书是集体智慧的结晶。北京工业大学计算机学院的刘波、胡海鹤和刘兆英等老师，以及张亚红、曾少锋、沈成恺、杨红丽和丁勇等同学，在文献和软件资料的收集整理方面提供了很大帮助。此外，华章公司的温莉芳副总经理对本书的出版给予了大力支持，张梦玲编辑对本书内容的编排提出了许多宝贵意见。在这里向他们表示衷心的感谢。

最后，还要感谢父母、爱人和儿女在本书写作期间给予的理解，感谢他们的真情鼓励、默默付出以及对非规律生活的宽容。同时，作者在此也因减少了对他们的关爱而深表愧疚和歉意。

限于作者水平，本书在内容取材和结构编排上可能存在不妥之处，希望使用本书的教师、学生、专家以及其他读者提出宝贵的批评和建议。

作者

2016 年 8 月于北京工业大学

目 录

前言

第一部分 基础理论

第1章 概述	2
1.1 深度学习的起源和发展	2
1.2 深层网络的特点和优势	4
1.3 深度学习的模型和算法	7
第2章 预备知识	9
2.1 矩阵运算	9
2.2 概率论的基本概念	11
2.2.1 概率的定义和性质	11
2.2.2 随机变量和概率密度函数	12
2.2.3 期望和方差	13
2.3 信息论的基本概念	14
2.4 概率图模型的基本概念	15
2.5 概率有向图模型	16
2.6 概率无向图模型	20
2.7 部分有向无圈图模型	22
2.8 条件随机场	24
2.9 马尔可夫链	26
2.10 概率图模型的学习	28
2.11 概率图模型的推理	29
2.12 马尔可夫链蒙特卡罗方法	31
2.13 玻耳兹曼机的学习	32
2.14 通用反向传播算法	35

2.15 通用逼近定理	37
-------------------	----

第3章 受限玻耳兹曼机

3.1 受限玻耳兹曼机的标准模型	38
3.2 受限玻耳兹曼机的学习算法	40
3.3 受限玻耳兹曼机的变种模型	44

第4章 自编码器

4.1 自编码器的标准模型	48
4.2 自编码器的学习算法	50
4.3 自编码器的变种模型	53

第5章 深层信念网络

5.1 深层信念网络的标准模型	57
5.2 深层信念网络的生成学习算法	60
5.3 深层信念网络的判别学习算法	62
5.4 深层信念网络的变种模型	63

第6章 深层玻耳兹曼机

6.1 深层玻耳兹曼机的标准模型	64
6.2 深层玻耳兹曼机的生成学习算法	65
6.3 深层玻耳兹曼机的判别学习算法	69
6.4 深层玻耳兹曼机的变种模型	69

第7章 和积网络

7.1 和积网络的标准模型	72
7.2 和积网络的学习算法	74

7.3 和积网络的变种模型	77	14.2 实验数据	118
第8章 卷积神经网络	78	14.3 代码工具	120
8.1 卷积神经网络的标准模型	78	第15章 自编码器降维案例	121
8.2 卷积神经网络的学习算法	81	15.1 自编码器降维程序的模块 简介	121
8.3 卷积神经网络的变种模型	83	15.2 自编码器降维程序的运行 过程	122
第9章 深层堆叠网络	86	15.3 自编码器降维程序的代码 分析	127
9.1 深层堆叠网络的标准模型	86	15.3.1 关键模块或函数的主要 功能	127
9.2 深层堆叠网络的学习算法	87	15.3.2 主要代码分析及注释	128
9.3 深层堆叠网络的变种模型	88	15.4 自编码器降维程序的使用 技巧	138
第10章 循环神经网络	89	第16章 深层感知器识别案例	139
10.1 循环神经网络的标准模型	89	16.1 深层感知器识别程序的模块 简介	139
10.2 循环神经网络的学习算法	91	16.2 深层感知器识别程序的运行 过程	140
10.3 循环神经网络的变种模型	92	16.3 深层感知器识别程序的代码 分析	143
第11章 长短时记忆网络	94	16.3.1 关键模块或函数的主要 功能	143
11.1 长短时记忆网络的标准模型	94	16.3.2 主要代码分析及注释	143
11.2 长短时记忆网络的学习算法	96	16.4 深层感知器识别程序的使用 技巧	148
11.3 长短时记忆网络的变种模型	98	第17章 深层信念网络生成 案例	149
第12章 深度学习的混合模型、 广泛应用和开发工具	102	17.1 深层信念网络生成程序的模块 简介	149
12.1 深度学习的混合模型	102	17.2 深层信念网络生成程序的运行 过程	150
12.2 深度学习的广泛应用	104	17.3 深层信念网络生成程序的代码 分析	153
12.2.1 图像和视频处理	104		
12.2.2 语音和音频处理	106		
12.2.3 自然语言处理	108		
12.2.4 其他应用	109		
12.3 深度学习的开发工具	110		
第13章 深度学习的总结、 批评和展望	114		
 第二部分 案例分析			
第14章 实验背景	118		
14.1 运行环境	118		

17.3.1	关键模块或函数的主要功能	153
17.3.2	主要代码分析及注释	153
17.4	深层信念网络生成程序的使用技巧	162
第 18 章 深层信念网络分类		
	案例	163
18.1	深层信念网络分类程序的模块简介	163
18.2	深层信念网络分类程序的运行过程	165
18.3	深层信念网络分类程序的代码分析	169
18.3.1	关键模块或函数的主要功能	169
18.3.2	主要代码分析及注释	170
18.4	深层信念网络分类程序的使用技巧	201
第 19 章 深层玻耳兹曼机识别		
	案例	202
19.1	深层玻耳兹曼机识别程序的模块简介	202
19.2	深层玻耳兹曼机识别程序的运行过程	203
19.3	深层玻耳兹曼机识别程序的代码分析	206
19.3.1	关键模块或函数的主要功能	206
19.3.2	主要代码分析及注释	206
19.4	深层玻耳兹曼机识别程序的使用技巧	220

第 20 章 卷积神经网络识别		
	案例	221
20.1	DeepLearnToolbox 程序的模块简介	221
20.2	DeepLearnToolbox 程序的运行过程	221
20.3	DeepLearnToolbox 程序的代码分析	223
20.3.1	关键函数的主要功能	223
20.3.2	主要代码分析及注释	223
20.4	DeepLearnToolbox 程序的使用技巧	227
20.5	Caffe 程序的模块简介	227
20.6	Caffe 程序的运行过程	228
20.7	Caffe 程序的代码分析	230
20.7.1	关键函数的主要功能	230
20.7.2	主要代码分析及注释	231
20.8	Caffe 程序的使用技巧	235
第 21 章 循环神经网络填充		
	案例	236
21.1	槽值填充的含义	236
21.2	循环神经网络填充程序的模块简介	236
21.3	循环神经网络填充程序的运行过程	237
21.4	循环神经网络填充程序的代码分析	238
21.4.1	关键函数的主要功能	238
21.4.2	主要代码分析及注释	238

21.5	循环神经网络填充程序的 使用技巧	244	22.3.1	关键模块或函数的 主要功能	247
第 22 章 长短时记忆网络分类			22.3.2 主要代码分析及 注释		
	案例	245	22.4 长短时记忆网络分类程序的 使用技巧		
22.1	长短时记忆网络分类程序的 模块简介	245	附录 1 Caffe 在 Windows 上的 安装过程		
22.2	长短时记忆网络分类程序的 运行过程	246	附录 2 Theano 的安装过程		
22.3	长短时记忆网络分类程序的 代码分析	247	参考文献		
			269		

PART I
第一部分

基础理论

本书第一部分主要探讨深度学习的基础理论。深度学习起源于神经网络，其本质是一系列深层网络模型的学习和训练算法。本部分涵盖了深度学习的主要内容，有助于读者在总体上把握深度学习的发展脉络和体系结构，是开展进一步相关工作的基础。

这部分共包括 13 章。第 1 章勾画深度学习的起源和发展、特点和优势、模型和算法。第 2 章介绍预备知识，读者可跳过熟悉的部分，但建议认真学习概率图模型、玻耳兹曼机和通用反向传播算法等难点内容，因为这些内容是理解许多深度学习模型和算法的基础。第 3~11 章，依次介绍深度学习的 9 种重要模型，包括受限玻耳兹曼机、自编码器、深层信念网络、深层玻耳兹曼机、和积网络、卷积神经网络、深层堆叠网络、循环神经网络、长短时记忆网络，而且对于其中的每一个模型，都从标准模型、学习算法和变种模型三个方面进行介绍。第 12 章讨论深度学习的若干混合模型、多种多样的应用以及常用的开源库。第 13 章总结深度学习的研究现状，明确存在的问题，并指出其未来的发展方向。

概 述

如何让机器从经验中学习长期以来都是哲学界和科学界的研究目标之一。学习能力对人类智能的形成和发展无疑起着至关重要的作用，而机器学习的研究显然有助于提高人工智能的水平。从原始的输入数据到产生意义的理解过程往往需要经过许多不同层次的信息处理、转换、表达和抽象，如果涉及的层次较深，深度学习的模型和方法就可能发挥重要作用。本章主要勾画深度学习的起源和发展、特点和优势、模型和算法。

1.1 深度学习的起源和发展

作为一种实现人工智能的强大技术，深度学习（deep learning）已经在手写数字识别、维数约简、语音识别、图像理解、机器翻译、蛋白结构预测和情感识别等各个方面获得了广泛应用^[1-7]，因屡屡取得打破记录的评测结果并超越其他方法，而很快受到了非常高度的关注。

深度学习的概念起源于人工神经网络（artificial neural network），在本质上是指一类对具有深层结构（deep architecture）的神经网络进行有效训练的方法。神经网络是一种由许多非线性计算单元（或称神经元、节点）组成的分层系统，通常网络的深度就是其中的不包括输入层的层数。理论上，一个具有浅层结构（shallow architecture）或层数不够深的神经网络虽然在节点数足够大时，也可能充分逼近地表达任意的多元非线性函数，但这种浅层表达在具体实现时往往由于需要太多的节点而无法实际应用。一般说来，对于给定数目的训练样本，如果缺乏其他先验知识，人们更期望使用少量的计算单元来建立目标函数的“紧表达”（compact representation），以获得更好的泛化能力^[8]。而在网络深度不够时，这种紧表达可能是根本无法建立起来的，因为理论研究表明，深度为 k 的网络能够紧表达的函数在用深度为 $k-1$ 的网络来表达时，有时需要的计算单元会指数增长^[9]。

最早的神经网络是心理学家 McCulloch 和数理逻辑学家 Pitts 在 1943 建立的 MP 模型^[10]，而相关的线性回归方法甚至可以追溯到 1800 年前后^[11]。MP 模型实际只是单个神经元的形式化数学描述，具有执行逻辑运算的功能，虽然不能进行学习，但开创了人工神经网络研究的时代。

1949年, Hebb 首先对生物神经网络提出了有关学习的思想^[12]。1958年, Rosenblatt 提出了感知器(perceptron)模型及其学习算法^[13]。在随后的几十年间, 尽管神经网络的研究出现过一段与 Minsky 对感知器的批评有关的低潮期^[14], 但 Grossberg^[15]、Kohonen^[16]、Narendra&Thathathchar^[17]、von der Malsburg^[18]、Widrow&Hoff^[19]、Palm^[20]、Willshaw&von der Malsburg^[21]、Hopfield^[22]、Ackley^[23]、Rumelhart^[24]等人仍然逐步提出了许多神经网络的新模型。到20世纪八九十年代, 这些新模型终于引发了神经网络的重生, 并掀起了对神经网络研究的世界性高潮^[25]。其中最受欢迎的模型至少包括: Hopfield 神经网络^[22]、玻耳兹曼机^[23]和多层感知器(Multilayer Perception, MLP)^[24]。最早的深度学习系统也许就是那些通过数据分组处理方法训练的多层感知器^[26]。多层感知器(在隐含层数大于1时又称为深层感知器)实际上是一种由多层节点有向图构成的前馈神经网络(Feedforward Neural Network, FNN)^[27], 其中每一个非输入节点是具有非线性激活函数的神经元, 每一层与其下一层是全连接的。此外, Fukushima 提出的神经认知机可能是第一个具有“深度”属性的神经网络^[28-31], 并且也是第一个集成简单细胞和复杂细胞的神经生理学洞见的神经网络, 以便有效地对视觉输入的某些特性起反应^[31,32]。更重要的是, 神经认知机促成了卷积神经网络结构的诞生和发展^[33]。而卷积神经网络作为一种判别模型, 在近几年的大规模数据评测比赛中成绩卓著^[34], 盛誉非凡。

在训练神经网络方面, 反向传播(backpropagation)无疑是最常用、最著名的算法, 最先由 Werbos 描述^[35], 由 LeCun^[36]和 Parker^[37]的有关论文发表, 由 Rumelhart 等人^[38]的论文享誉全球, 并由 Battiti^[39]、Fahlman^[40]、Igel&Hüsken^[41]、Jacobs^[42]、Neuneier 和 Zimmermann^[43]、Orr&Müller^[44]、Riedmiller&Braun^[45]、Schraudolph&Sejnowski^[46]、West&Saad^[47]等人的工作得到了更好的发展。然而, 直到20世纪80年代晚期, 反向传播似乎还只是对浅层网络有效, 尽管原理上也应对深层网络有效。事实上, 大多数多层感知器的应用都只用到很少的隐含层, 增加隐含层几乎没有什么经验上的收益。这似乎可以从神经网络的逼近定理中找到某种解释^[48,49], 该定理为: 只要单隐层感知器包含的隐含神经元足够多, 就能够在闭区间上以任意精度逼近任何一个多变量连续函数。直到1991年时, 这个关于多层感知器在增加层数时为什么难学的问题, 才开始作为一个深度学习的基本问题, 得到了完全的理解, 之前的其他可能想法变得不再重要。

1991年, Hochreiter 正式指出, 典型的深层网络存在梯度消失或爆炸问题(the problem of vanishing or exploding gradient), 从而明确确立了深度学习的一个里程碑^[50]。该问题为: 累积反向传播误差信号在神经网络的层数增加时会出现指数衰减或增长的现象, 从而导致数值计算快速收缩或越界。这就是为什么深层网络很难用反向传播算法训练的主要原因。需要指出的是, 梯度消失或爆炸问题又称为长时滞后问题(the long time lag problem), 在循环神经网络中也会出现^[51]。

为了在一定程度上克服梯度消失或爆炸问题, 从1990到2000年前后, Hochreiter 的深邃思

想推动了若干新方法的探索,包括:1991 年的一个非常深的网络^[52]、长短时记忆网络(long short-term memory network)^[53]、基于 GPU 的计算机^[54]、海森无关优化(Hessian-free optimization)^[55]、权值矩阵空间的替代搜索(如随机权值猜测^[56]、通用搜索^[57]、evolino^[58]和直接搜索^[59])。但除了卷积神经网络以外^[60],训练深层网络的问题直到 2006 年才开始受到严肃认真的对待。一个重要的原因是 1995 年之后支持向量机的快速发展^[61],让神经网络的有关工作黯然失色。

普遍认为,深度学习正式发端于 2006 年,以 Hinton 及其合作者发表的两篇重要论文为标志,一篇发表在《Neural Computation》上,题目为“A fast learning algorithm for deep belief nets”^[62],另一篇发表在《Science》上,题目为“Reducing the dimensionality of data with neural networks”^[1]。从那时起,深度学习的概念就开始明朗,因为清楚地说明了浅层网络的无监督学习(如受限玻耳兹曼机^[63]的对比散度算法^[64])有助于深层网络(如深层自编码器和深层感知器^[1])的有监督学习算法(如反向传播)。特别是,经验证实了由无监督预训练和有监督调优构成的两阶段策略不仅对于克服深层网络的训练困难是有效的,而且赋予了深层网络优越的特征学习能力。紧接着,许多新的深层结构又被建立起来(如深层玻耳兹曼机^[65]、和积网络^[66]、深层堆叠网络^[67]),许多新的可行技术被发展起来(如最大池化^[68]、丢失连接(dropconnect)^[69]、dropout^[70]),而且还取得了许多新的历史性成就(如手写体数字识别^[1]、ImageNet 分类^[3]和语音识别^[71])。所有这些结构、技术和成就,使深度学习大受欢迎,并很快在学术界发展成为一次神经网络的新浪潮。这其中最主要的原因当然是深度学习在大量重要应用中的性能超越了机器学习的其他替代方法(如支持向量机^[72])。

深度学习的巨大成功也迅速点燃了一大批杰出学者的热情。随着研究的不断发展,各国政府部门逐步给深度学习投入了大量的科研经费,在 ICML、NIPS、ICLR、IEEE Trans. PAMI 等著名会议和期刊上发表的相关论文越来越多,而且深度学习频频打破评测记录的战果还极大地激发了工业界的斗志,Google、Facebook、微软、百度、Apple、IBM、Yahoo!、Twitter、腾讯、京东和阿里巴巴,以及许多其他搜索和社交公司,都加入了其中的比拼,有的甚至为之疯狂。从目前的情况看,这场声势浩大的深度学习浪潮至少还要持续一段相当长的时间。

1.2 深层网络的特点和优势

神经网络由许多简单的、互连的称为神经元的处理器组成。每一个神经元产生一系列的实值激活^[73],其中输入神经元通过传感器激活,其余神经元通过连接激活。

例如,图 1.1 是两个浅层网络的例子,其中图 1.1a 是一个单隐层的普通神经网络,图 1.1b 是一个单隐层的和积网络。图 1.2 是两个深层网络的例子,其中图 1.2a 是一个多层神经网络,图 1.2b 是一个多层和积网络。

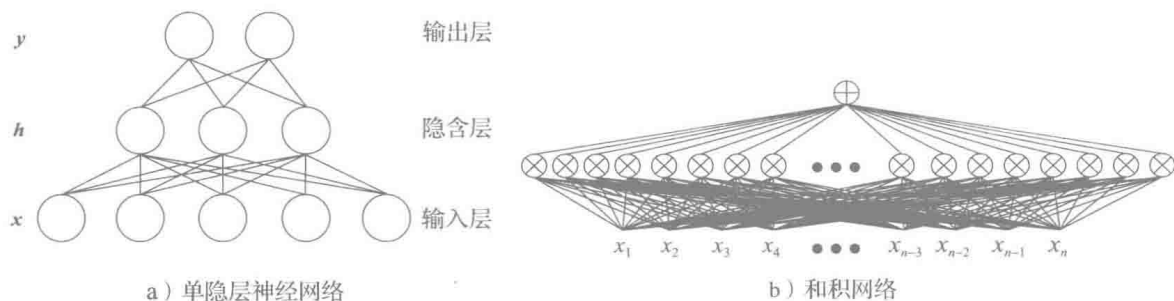


图 1.1 浅层网络举例

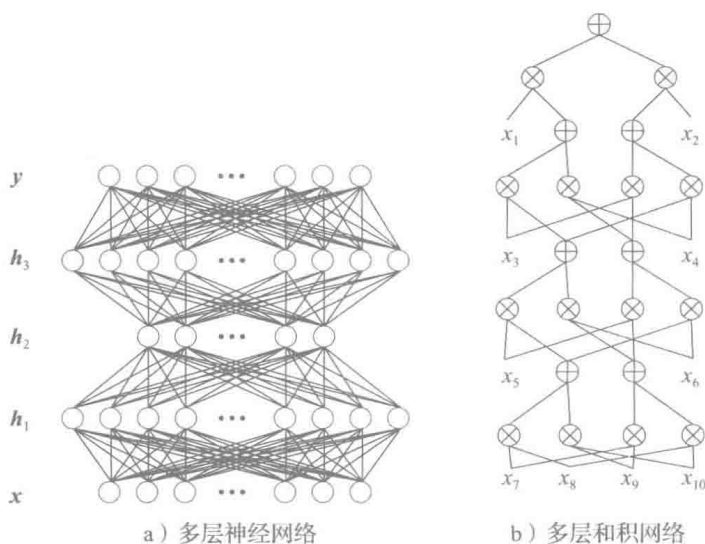


图 1.2 深层网络举例

根据 Bengio 的定义^[8]，深层网络由多层自适应非线性单元组成。换句话说，深层网络是非线性模块的级联，在所有层次上都包含可训练的参数。在理论上，深层网络和浅层网络的数学描述是类似的，而且都能够通过函数逼近表达数据的内在关系和本质特征。不过应注意，网络虽然在狭义上是指由神经元构成的神经网络，但在广义上可以指任何具有网络结构的学习模型。

迄今还没有公认的区分深层网络和浅层网络的深度划界标准。依据 Schmidhuber 的观点^[73]，深层网络和浅层网络可以用得分路径（或译为信度分配路径，Credit Assignment Path, CAP）深度加以区分。得分路径是一条可学习的、连接行为和结果的因果链。对于前馈神经网络，得分路径深度，也就是网络深度，是网络的隐含层数加 1（输出层也是可学习的）。对于循环神经网络，得分路径长度可能是无限的，因为信号可以多次通过同一个层。一般认为深层网络至少包含 3 个非输入层或者 $CAP > 2$ ，而非常深的网络应该深度（或 CAP）至少大于 10。在工程实践中，深层网络通常是一个多层人工神经网络，可以包含多个隐含层和多达几百万个自由参数。

浅层网络对机器学习来说也很重要,包括单隐层网络^[74]、高斯混合模型 (Gaussian Mixture Model, GMM)^[75]、隐马尔可夫模型 (Hidden Markov Model, HMM)^[76]、条件随机场 (Conditional Random Field, CRF)^[77]、支持向量机 (Support Vector Machine, SVM)^[78]、逻辑回归^[79]、最大熵模型^[80],等等。这些网络的共同特点是,它们都使用不超过三层的结构将原始输入信号变换到一个特征空间。毋庸置疑,浅层网络对解决许多简单的和有良好约束的问题非常有效,但在解决真实世界的复杂应用问题时,往往出现函数表达能力不足的情况。这是因为在处理某些问题时,可能需要指数增长的计算单元,而此时深层网络则可能仅需相对很少的计算单元^[81]。

作为例子,不妨来分析一个具有递归结构的和积网络的函数表达能力。设输入变量的个数 $n=4^i$, 其中 i 是正整数。 l^0 代表输入层, 其中第 j 个节点表示为 $l_j^0 = x_j$, $1 \leq j \leq n$ 。分别构造奇数层和偶数层的节点如下:

$$\begin{cases} l_j^{2k+1} = l_{2j-1}^{2k} \cdot l_{2j}^{2k}, & 0 \leq k \leq i-1 \quad \text{和} \quad 1 \leq j \leq 2^{2(i-k)-1} \\ l_j^{2k} = \lambda_{jk} l_{2j-1}^{2k-1} + \mu_{jk} l_{2j}^{2k-1}, & 1 \leq k \leq i \quad \text{和} \quad 1 \leq j \leq 2^{2(i-k)} \end{cases} \quad (1.1)$$

其中,权值 λ_{jk} 和权值 μ_{jk} 都为正数。

该和积网络的输出 $f(x_1, \dots, x_n) = l_1^{2i} \in \mathbf{R}$ 是一个单节点。当 $i=1$ 时,网络共有 3 个非输入节点,结构如图 1.3 所示。由于对任意正整数 i , 这个和积网络在不计输入层时共有 $2i$ 层, 其中包含的(非输入)节点总数为 $1 + 2 + 4 + 8 + \dots + 2^{2i-1} = 2^{2i} - 1 = 4^i - 1 = n - 1$, 所以网络规模仅具有线性复杂度。显然,这个递归和积网络在 $i > 1$ 时是一个深层网络。

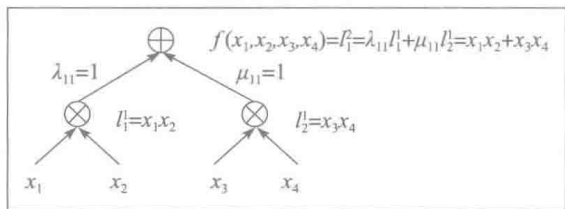


图 1.3 输入 $n=4$ 的和积网络

如果用图 1.1b 中的单隐层和积网络来计算函数 $f(x_1, x_2, \dots, x_n)$, 那么需要把它改写成输入变量乘积的加权和形式。当所有权值都取 1 时,可以得到下面的表达式:

$$f(x_1, x_2, \dots, x_{4^i}) = x_1 x_2 x_3 x_4 \dots x_{4^{i-1}-3} x_{4^{i-1}-2} + \dots \quad (1.2)$$

由于在该表达式中乘积项的数量为 $m_{2i} = 2^{\sqrt{n}-1}$, 因此用单隐层和积网络计算需要 $2^{\sqrt{n}-1}$ 个积节点和一个和节点, 共需 $2^{\sqrt{n}-1} + 1$ 个节点, 网络规模具有指数复杂度。因为在 n 较大时, $2^{\sqrt{n}-1} + 1$ 将远远大于 $n-1$, 所以用浅层和积网络计算具有 n 个输入的函数, 需要的节点个数可能比深层和积网络多得多。例如, 当 $n=4^5 = 1024$ 时, 用浅层和积网络计算 $f(x_1, \dots, x_n) = l_1^{2i}$, 需要 $2^{\sqrt{1024}-1} + 1 = 2^{31} + 1 = 2147483649$ 个节点, 而用深层和积网络仅需 $1024 - 1 = 1023$ 个节点。

由此可见,在表达同样的复杂函数时,与浅层网络相比,深层网络可能只需要很少的节点和很少的参数。这意味着,在总节点数大致相同的情况下,深层网络通常比浅层网络的函数表达能力更强。

1.3 深度学习的模型和算法

深度学习亦称深度机器学习、深度结构学习、分层学习，是一类有效训练深层神经网络（Deep Neural Network, DNN）的机器学习算法，可以用于对数据进行高层抽象建模。广义上说，深层神经网络是一种具有多个处理层的复杂结构，其中包含多重非线性变换。如果深度足够，那么多层感知器无疑是深层网络，前馈神经网络也是深层网络。基本的深层网络模型可以分为两大类：生成模型和判别模型。生成是指从隐含层到输入数据的重构过程，而判别是指从输入数据到隐含层的归约过程。复杂的深层结构可能是一个混合模型，既包含生成模型成分，又包含判别模型成分。生成模型一般用来表达数据的高阶相关性或者描述数据的联合统计分布，判别模型则通常用来分类数据的内在模式或者描述数据的后验分布。生成模型主要包括受限玻耳兹曼机（Restricted Boltzmann Machine, RBM）、自编码器（Autoencoder, AE）^[1]、深层信念网络（Deep Belief Network, DBN）^[62]、深层玻耳兹曼机（Deep Boltzmann Machine, DBM）^[65]以及和积网络（Sum-Product Network, SPN）^[66]，其中 AE、DBN 和 DBM 需要 RBM 进行预训练。判别模型主要包括深层感知器（deep MLP）、深层前馈网络（deep FNN）、卷积神经网络（Convolutional Neural Network, CNN）^[82]、深层堆叠网络（Deep Stacking Network, DSN）^[83]、循环神经网络（Recurrent Neural Network, RNN）^[84]和长短时记忆（Long Short-Term Memory, LSTM）网络^[53]。值得一提的是，虽然受限玻耳兹曼机、自编码器、深层信念网络、深层玻耳兹曼机，以及和积网络都被归类为生成模型，但由于模型中也包含判别过程（即从输入到隐含层的规约），所以在一定条件下，也可以看作判别模型并用于对数据的分类和识别，而且在用于产生序列数据时，循环神经网络也可以看作是生成模型。此外，虽然受限玻耳兹曼机作为一种两层网络，在严格意义上并不是一种深层网络，但由于它是对许多深层网络进行预训练的基础，所以也被看作一种基本的深度学习模型。自编码器作为一种深度学习模型，通常只是用作其他模型的构建模块，而不是作为一个独立的模型使用。

基于各种模型和算法，深层网络能够从大量的复杂数据中学习到合适且有效的特征。这些特征在解决实际问题时常常能够取得极佳的效果，从而使得深度学习受到了学术界和工业界的普遍青睐。借助无监督学习，前馈神经网络和循环神经网络的纯有监督学习早已在有关评测比赛中崭露头角^[85,86]，在大多数近年的比赛中更是成绩卓著^[87-94]。特别地，基于 GPU 的最大池化卷积神经网络，不仅在模式识别和图像分类的比赛中捷报频传^[95-97]，而且在图像分割和目标检测的比赛中也是战果累累、名列前茅^[3,96,99]。目前，基于深度学习的机器玩家，通过结合卷积神经网络和强化学习，只需输入图像像素和游戏分数进行训练，就能够学会有效的操作策略，在很多视频游戏中达到与人类专业玩家相当的水平^[100]。而最为空前的是，一个命名为 AlphaGo 的人工智能程序机器人，利用深层网络和蒙特卡罗树搜索（Monte Carlo tree search），

首次在完整的围棋比赛中战胜了人类的专业选手、欧洲冠军、职业围棋二段选手樊麾，没有任何让子，且以 5 比 0 获胜^[101]。这在围棋人工智能领域，是一次史无前例的突破。而且，在 2016 年 3 月，AlphaGo 又以 4 比 1 战胜了人类的顶尖高手、世界冠军、职业围棋九段选手李世石（或李世乭），这更是一次亘古未有的创举。

本书的主要内容就是以矩阵运算、概率论、信息论、图模型等预备知识为基础，从受限玻耳兹曼机开始，逐一介绍深度学习的主要模型和算法，包括它们的变种模型和混合模型，以及广泛应用和问题挑战，并分析一些基本案例。