

Introduction to Deep Learning Using R

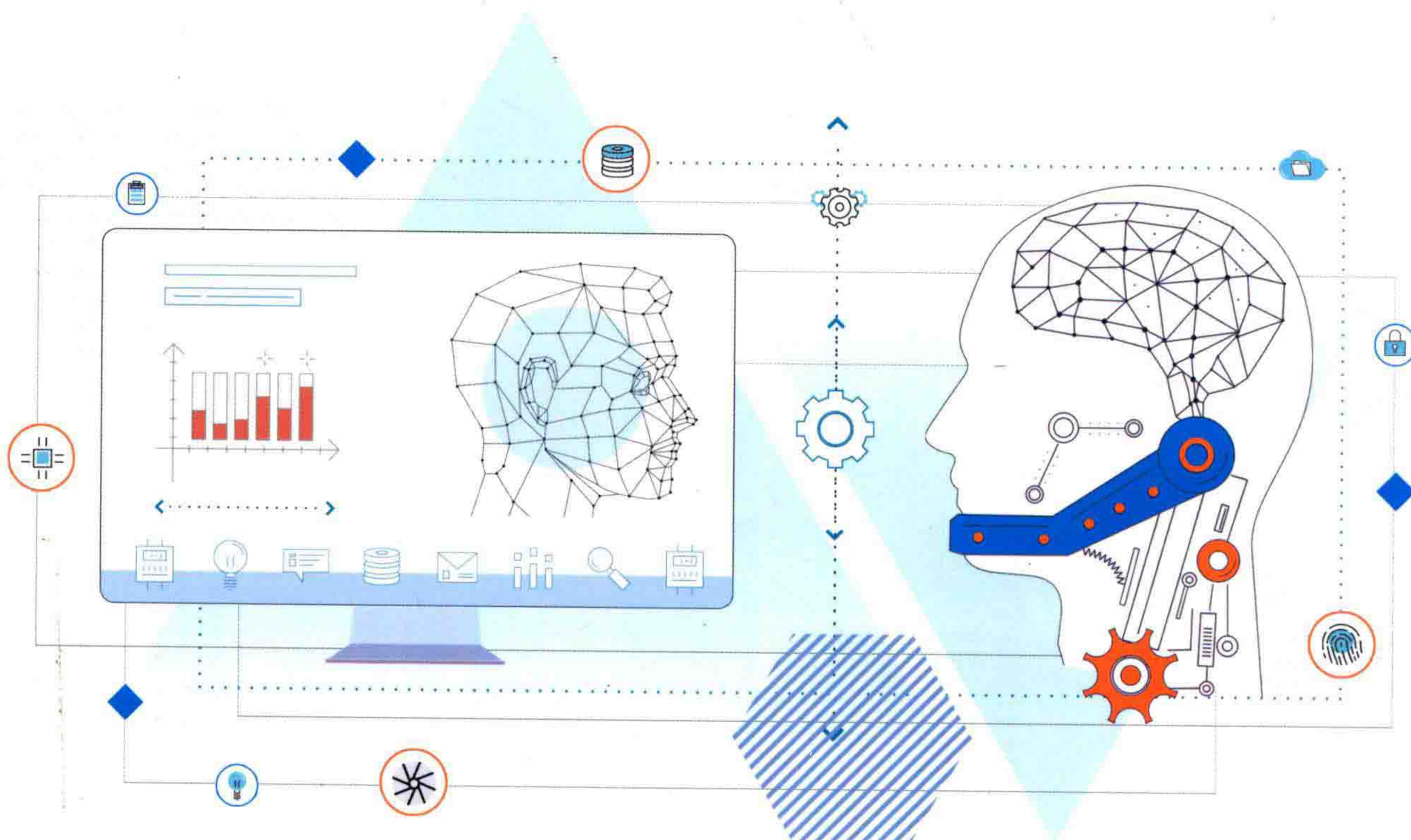
A Step-by-Step Guide to Learning and
Implementing Deep Learning Models Using R

深度学习

R语言实践指南

[美] 托威赫·贝索洛 (Taweh Beysolow II) 著

潘怡 译



机械工业出版社
China Machine Press

Introduction to Deep Learning Using R

A Step-by-Step Guide to Learning and
Implementing Deep Learning Models Using R

深度学习

R语言实践指南

[美] 托威赫·贝索洛 (Taweh Beysolow II) 著

潘怡 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

深度学习: R 语言实践指南 / (美) 托威赫·贝索洛 (Taweh Beysolow II) 著; 潘怡译.
—北京: 机械工业出版社, 2018.7

(智能系统与技术丛书)

书名原文: Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning
and Implementing Deep Learning Models Using R

ISBN 978-7-111-60437-2

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

中国版本图书馆 CIP 数据核字 (2018) 第 156147 号

本书版权登记号: 图字 01-2017-8952

Taweh Beysolow II: Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and
Implementing Deep Learning Models Using R (ISBN: 978-1-4842-2733-6).

Original English language edition published by Apress Media.

Copyright © 2017 by Taweh Beysolow II. Simplified Chinese-language edition copyright © 2018 by
China Machine Press. All rights reserved.

This edition is licensed for distribution and sale in the People's Republic of China only, excluding
Hong Kong, Taiwan and Macao and may not be distributed and sold elsewhere.

本书原版由 Apress 出版社出版。

本书简体字中文版由 Apress 出版社授权机械工业出版社独家出版。未经出版者预先书面许可, 不得以任何方式
复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售发行, 未经授权的本书出口
将被视为违反版权法的行为。

深度学习: R 语言实践指南

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: 殷虹

印刷: 北京诚信伟业印刷有限公司

版次: 2018 年 7 月第 1 版第 1 次印刷

开本: 186mm × 240mm 1/16

印张: 14

书号: ISBN 978-7-111-60437-2

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) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@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

华章IT
HZBOOKS | Information Technology



THE TRANSLATOR'S WORDS

译者序

2013 年,《MIT Technology Review》杂志将深度学习列为年度十大突破性技术之首。作为当下热门的科技领域之一,深度学习在诸如无人驾驶、社交平台上的图像识别、智能翻译等领域相继取得了突破性的进展,无论是在工业界还是在学术界均引起了广泛关注,并源源不断地涌现出了许多丰富而有益的创造。但不可否认的是,由于对该领域知识的匮乏,即使是专家也可能在特定情况下因使用一些比较差劲的工具而得到糟糕的结果,更不用说深度学习的初学者。因此,本书的初衷是通过对机器学习、深度学习基本概念和基础理论的介绍及实际案例的讲解,帮助读者了解相关技术的用途,并能有效地将它们应用于各自的工作中。

本书共 11 章,内容主要涉及:深度学习的数学理论基础,包括重要的统计学和线性代数的相关基本概念和知识;深度学习的各种典型模型,例如传统的单层感知器模型、多层感知器模型,以及卷积神经网络、循环神经网络、受限玻耳兹曼机、深度信念网络等一些更为复杂的模型;构建深度学习模型的实验设计方法以及实验过程中的特征选择方法;应用 R 语言进行机器学习和深度学习实践的案例,通过这些案例,读者可以增强学习机器学习和深度学习的信心,提高对实际问题的解决能力。

感谢翻译过程中华章公司张梦玲编辑的鼓励与督促,感谢家人和朋友们一如既往的支持和鼓励。

尽管此前译者已经翻译完成了两本 R 语言与机器学习领域的专业书籍,但由于深度学习技术是一个快速发展的方向,新的模型和应用层出不穷,使得本书的翻译工作比预期要艰难许多,兼之本人学识有限以及中英文语言表达、术语翻译上的习惯,书中难免存在错误,还请广大读者指正与原谅。

ABOUT THE AUTHOR

关于作者

Taweh Beysolow II 机器学习科学家，现居美国，热衷于研究及应用机器学习方法解决实际问题。他本科毕业于圣约翰大学，获得经济学学士学位，后获得福特汉姆大学应用统计学硕士学位。他对一切与机器学习、数据科学、计量金融及经济学相关的内容都有着巨大的热情。

ABOUT THE TECHNICAL REVIEWER

关于技术审稿人



Somil Asthana 分别于 IITBHU 大学（印度）和布法罗大学（美国）获得计算机科学学士及硕士学位。他是一位企业家、机器学习魔术师以及大数据专家，为 Sprint、Verizon、HPE 和 Avaya 这类世界 500 强公司提供咨询服务。他拥有自己的创业公司，为电子商务、内容服务/媒体类数据驱动型公司提供大数据解决方案和数据策略。

P R E F A C E

前 言

本书假设读者已经掌握了统计学或计算机编程（特别是 R 语言）的基本知识。如果读者不具备这些基础，则有可能在代码实现部分遇到困难，这时建议回头重新学习以上提到的基础知识。

致谢

感谢我的家人。感谢我的祖母，我从她那里获益良多。也感谢本书的编辑和其他在完成这本书的过程中给予我支持的专家。感谢我的老师，感谢他们给予我源源不断的激励，使我一直保持着对研究的好奇心。感谢我的朋友，不论旧识、新交，他们让我的生活更有价值，也让我拥有了许多难忘的人生回忆。感谢我已故的朋友 Michael Giangrosso，我曾希望和他一起踏上深度学习的研究之旅。最后，感谢我已故的导师和朋友 Lawrence Sobol，永远感激他的指引，在我生命的每一天，都将铭记他的教诲。

CONTENTS

目 录

译者序		第2章 数学知识回顾	10
关于作者		2.1 统计学基本概念	10
关于技术审稿人		2.1.1 概率	10
前言		2.1.2 交与并	11
第1章 深度学习简介.....	1	2.1.3 贝叶斯定理	13
1.1 深度学习模型	3	2.1.4 随机变量	13
1.1.1 单层感知器模型	3	2.1.5 方差	14
1.1.2 多层感知器模型	4	2.1.6 标准差	15
1.1.3 卷积神经网络	4	2.1.7 可决系数	16
1.1.4 循环神经网络	5	2.1.8 均方误差	16
1.1.5 受限玻耳兹曼机	6	2.2 线性代数	16
1.1.6 深度信念网络	6	2.2.1 标量和向量	16
1.2 其他	7	2.2.2 向量的特性	17
1.2.1 实验设计	7	2.2.3 公理	18
1.2.2 特征选择	7	2.2.4 子空间	19
1.2.3 机器学习及深度学习		2.2.5 矩阵	19
应用	7	2.3 小结	41
1.2.4 深度学习的历史	8	第3章 优化及机器学习回顾	42
1.3 小结	9	3.1 无约束优化	42
		3.1.1 局部极小值	44

3.1.2 全局极小值	44	3.14 强化学习	80
3.1.3 局部极小值的条件	45	3.15 小结	82
3.2 近邻算法	46	第4章 单层及多层感知器模型 ...	83
3.3 机器学习方法：有监督学习	47	4.1 单层感知器模型	83
3.3.1 机器学习的历史	47	4.1.1 训练感知器模型	84
3.3.2 什么是算法	48	4.1.2 WH 算法	84
3.4 回归模型	48	4.1.3 单层感知器模型的 局限性	85
3.5 选择合适的学习速率	52	4.1.4 汇总统计结果	87
3.5.1 牛顿法	57	4.2 多层感知器模型	88
3.5.2 Levenberg-Marquardt 启发式 方法	57	4.2.1 收敛得到全局最优解	88
3.6 多重共线性	58	4.2.2 MLP 模型中的反向传播 算法	89
3.7 评价回归模型	61	4.2.3 MLP 模型的局限性和 讨论	91
3.8 分类	62	4.2.4 应该使用几层隐含层，又 应该有多少个神经元	92
3.8.1 逻辑回归	62	4.3 小结	94
3.8.2 受试者工作特征曲线	63	第5章 卷积神经网络	95
3.8.3 混淆矩阵	64	5.1 CNN 的结构和特点	95
3.8.4 逻辑回归的局限性	65	5.2 CNN 的组成	97
3.8.5 支持向量机	66	5.2.1 卷积层	97
3.9 机器学习方法：无监督学习	69	5.2.2 池化层	99
3.9.1 K-均值聚类	70	5.2.3 修正线性单元层	100
3.9.2 K-均值聚类的局限性	70	5.2.4 全连接层	100
3.10 最大期望算法	72	5.2.5 损失层	100
3.11 决策树学习	73	5.3 参数调整	101
3.12 集成方法以及其他启发式 算法	77		
3.13 贝叶斯学习	78		

5.4	经典的 CNN 架构	102	8.7	A/B 测试	140
5.5	正则化	104	8.7.1	简单双样本 A/B 测试 ...	141
5.6	小结	105	8.7.2	A/B 测试中的 β 二项 层次模型	141
第 6 章	循环神经网络	106	8.8	特征、变量选择技术	143
6.1	完全循环网络	106	8.8.1	后向与前向选择	143
6.2	使用时间反向传播训练 RNN ...	107	8.8.2	主成分分析	144
6.3	Elman 神经网络	108	8.8.3	因子分析	146
6.4	神经历史压缩器	109	8.9	处理分类数据	147
6.5	长短期记忆网络	109	8.9.1	因子水平编码	148
6.6	RNN 里的结构化抑制	111	8.9.2	分类标签问题：太多 水平值	148
6.7	参数调优更新算法	112	8.9.3	典型相关分析	148
6.8	RNN 的实际案例：模式检测 ...	112	8.10	包裹式、过滤式及嵌入式 算法	149
6.9	小结	116	8.11	其他局部搜索算法	150
第 7 章	自编码器、受限玻耳兹 曼机及深度信念网络 ...	117	8.11.1	登山算法	150
7.1	自编码器	117	8.11.2	遗传算法	150
7.2	受限玻耳兹曼机	119	8.11.3	模拟退火	151
7.3	深度信念网络	126	8.11.4	蚁群优化算法	152
7.4	快速学习算法	126	8.11.5	变邻域搜索算法	153
7.5	小结	128	8.12	反应式搜索优化	154
第 8 章	实验设计与启发	129	8.12.1	反应式禁忌	154
8.1	方差分析	129	8.12.2	固定禁忌搜索	155
8.2	F 统计和 F 分布	130	8.12.3	反应式禁忌搜索	157
8.3	Plackett-Burman 设计	137	8.12.4	WalkSAT 算法	157
8.4	空间填充	138	8.12.5	K 近邻	158
8.5	全因子	139			
8.6	Halton、Faure 和 Sobol 序列 ...	139			

8.13 小结	158	10.1.4 模型评价	171
第9章 软硬件建议	159	10.2 问题2: 速配	175
9.1 使用标准硬件处理数据	159	10.2.1 问题类型: 分类	175
9.2 固态硬盘和硬盘驱动器	160	10.2.2 数据预处理: 数据清洗 和填充	176
9.3 图形处理单元	160	10.2.3 特征选择	179
9.4 中央处理器	161	10.2.4 模型训练和评价	180
9.5 随机存取存储器	162	10.3 小结	188
9.6 主板	162	第11章 深度学习及其他	
9.7 供电设备	163	实例	189
9.8 机器学习软件的优化	163	11.1 自编码器	189
9.9 小结	164	11.2 卷积神经网络	196
第10章 机器学习实例	165	11.2.1 预处理	197
10.1 问题1: 资产价格预测	165	11.2.2 模型构建和训练	200
10.1.1 问题类型: 有监督 学习——回归	166	11.3 协同过滤	206
10.1.2 实验说明	168	11.4 小结	210
10.1.3 特征选择	169	结束语	211

CHAPTER 1

第 1 章

深度学习简介

随着硬件技术的发展及大数据的涌现，先进计算方法的应用也日益普及，市场对更好产品的渴望以及公司对资源优化的需求都成为技术创新的核心推动力。为了应对不断增长的市场需求，越来越多的人重新将兴趣转移至机器学习，在一个由统计学、数学和计算机科学形成的交叉学科领域，机器学习强调利用设计使算法能够在使用过程中经过迭代而不断优化。机器学习最初源自人工智能技术，但受限于当时理论及技术水平的不足，机器学习一开始顺理成章地只注重在一些特殊任务上对这些人工智能算法的应用。直至目前，绝大多数机器学习算法也主要集中于函数优化，甚至有时候即使人们找到了问题的解决方案，也不一定能很好地解释数据的隐含模式以及取得人工智能技术一直希望获得的推理能力，因此对机器学习算法的实践往往成为一个反复试错的过程，通过不断地比较实验结果来选择最优算法。某些时候这样做是没问题的，但是在语言建模和计算机视觉等方面，试错的方法就行不通了。

作为当下理论和技术最重要的突破，深度学习很好地克服了机器学习的某些短板，因而迅速地普及应用，成为当下最热门的科技领域之一。在诸如无人驾驶、社交平台上的图像识别、智能翻译等领域都能看见深度学习的影子。深度学习（Deep learning）是机器学习领域的一个分支，相比传统机器学习算法，深度学习算法能够对数据完成更高或更低级别的抽象、解释和学习。深度学习模型源于交叉的多知识领域，例如博弈论和神经科学，很多模型都来自对人类神经系统基础结构的模仿。随着深度学习的

发展，许多研究人员已经开始期待未来人们能够针对各类问题给出一个更模糊、更普遍的软件解决方案而不是像现在这样使用麻烦的硬编码方法。

深度学习和机器学习的问题域在一开始的时候有相似之处，但机器学习的主要关注点是不同复杂度的约束可满足性问题，而深度学习对问题的定义更广泛一些，其算法能根据复杂问题的不同层次理解数据的多级表示，换句话说而言，也就是算法不但拥有预测和分类的能力，同时还具有对不同复杂度问题的自适应处理能力。例如，在图像识别里，神经网络算法可以识别出睫毛、面部以及人类等目标。而深度学习的意义也很显然，那就是依托深度学习，人们就能够构建出智能软件来解决足够复杂的问题。例如，具备自动纠错的特点、能够根据每个人的词汇量来构建合适的语音模型。

深度学习模型由许多非线性单元组成多层网络来处理数据，这些单元也称为神经元，不同层次的网络对数据的抽象程度不同，图 1-1 是对一个神经网络层次结构的可视化描述。

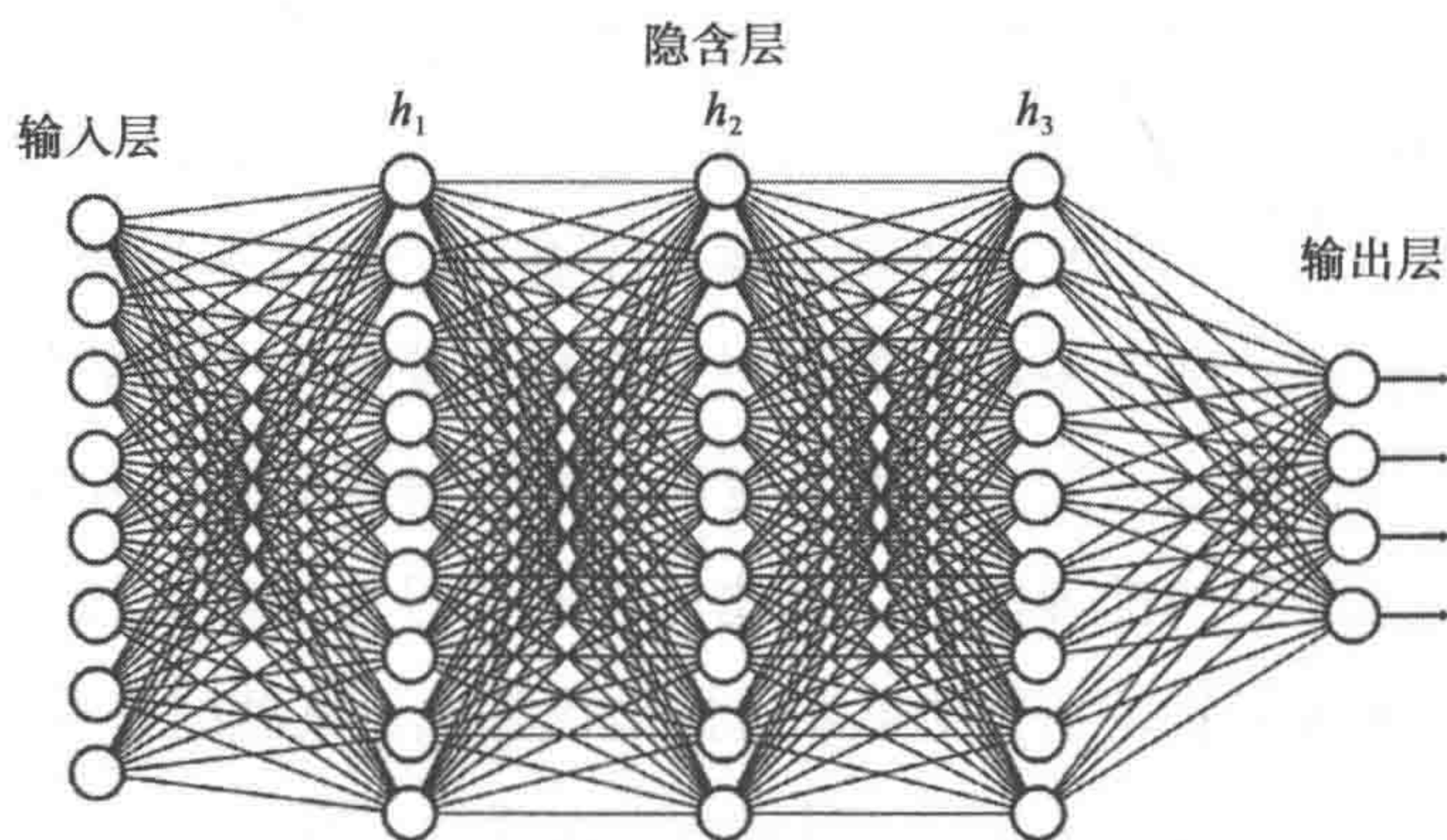


图 1-1 深度神经网络

深度神经网络因拥有多个隐含层而与传统神经网络区分开来，我们不需要确切知道这些“隐含”的网络层具体完成了哪些输入和输出功能，只需要清楚它们是上一层的输出即可。隐含层的叠加方式以及神经元之间的相互作用函数彼此都不相同，这也使得这些网络能够应对给定模型的不同应用需求。

更具体一点，就是说底层神经单元负责解释“方法”，而更高层的神经单元负责解

释“原理”。不同层使用的方法依据具体问题变化而变化，用户对方法调优后能取得更好的性能。相对更专注于分类和预测类问题的传统机器学习算法，深度神经网络算法的自适应性更强，这是因为深度神经网络理论源于人们假设现实世界中待解释的数据可以由不同层次的事实组合而成，多层网络结构确保了算法能够从简单的问题入手来构建对更复杂问题的分析模型。深度神经网络的诞生意味着人们不再需要像使用大多数传统机器学习方法一样对每个问题都给出同样明确的指示。而应用这些深度神经网络模型的基本优势之一就是它们处理无监督学习问题的能力，或者处理诸如给出一组响应变量 y 应该如何得到合适的解释变量 x 这样的难题的能力。以图像识别问题为例，选择一个已经由某个数据集训练后的模型，将一幅狗的图像加入到测试集中并假定模型不知道图像内容。算法执行过程中，会首先识别出狗的眼睫毛，接着再识别出狗的鼻子，然后是狗头部的轮廓，以此类推，直至识别出完整的狗的图像。

1.1 深度学习模型

现在，我们对深度学习已经有了一个大概的认识，接下来该具体了解一下本书到底想告诉你哪些有用的方法和模型了。

尽管本书假设读者已经掌握了基本的数学和统计学知识，但我们依然将线性代数、优化和机器学习的相关基础概念做简略介绍，以便读者能够牢固掌握深度学习的基础知识，更准确地理解深度学习技术。不过，数学基础好的读者也不要担心，本书对基础的背景知识介绍主要是考虑到如果部分读者确实对此有所需求，可以帮助他们能更好地开展下一步的研究。但本书最主要的目的还是探讨机器学习和深度学习模型的应用方法，而不是讲解冗长的理论知识。

当对必需的数学和机器学习概念有了足够的了解之后，我们将继续深入探讨机器学习模型。本章将介绍和解释深度学习模型。

1.1.1 单层感知器模型

单层感知器（Single Layer Perceptron, SLP）模型是最简单的神经网络模型，也是

其他高级深度学习模型的基础。通常情况下，在处理二元或多元分类问题时，可以使用 SLP 模型根据输入数据完成对象标注。在 SLP 模型中，输入层和输出层直接相连，每一个输入层和其上权值相乘，然后乘积之和加偏置，成为激励函数的输入。而最终感知器模型的输出结果则由激励函数的输出与某个用户给定的阈值比较结果决定。图 1-2展示了与 SLP 模型类似的麦卡洛克 - 皮茨神经模型（McCulloch- Pitts Neuron），该模型由研究者于 20 世纪 40 年代提出。

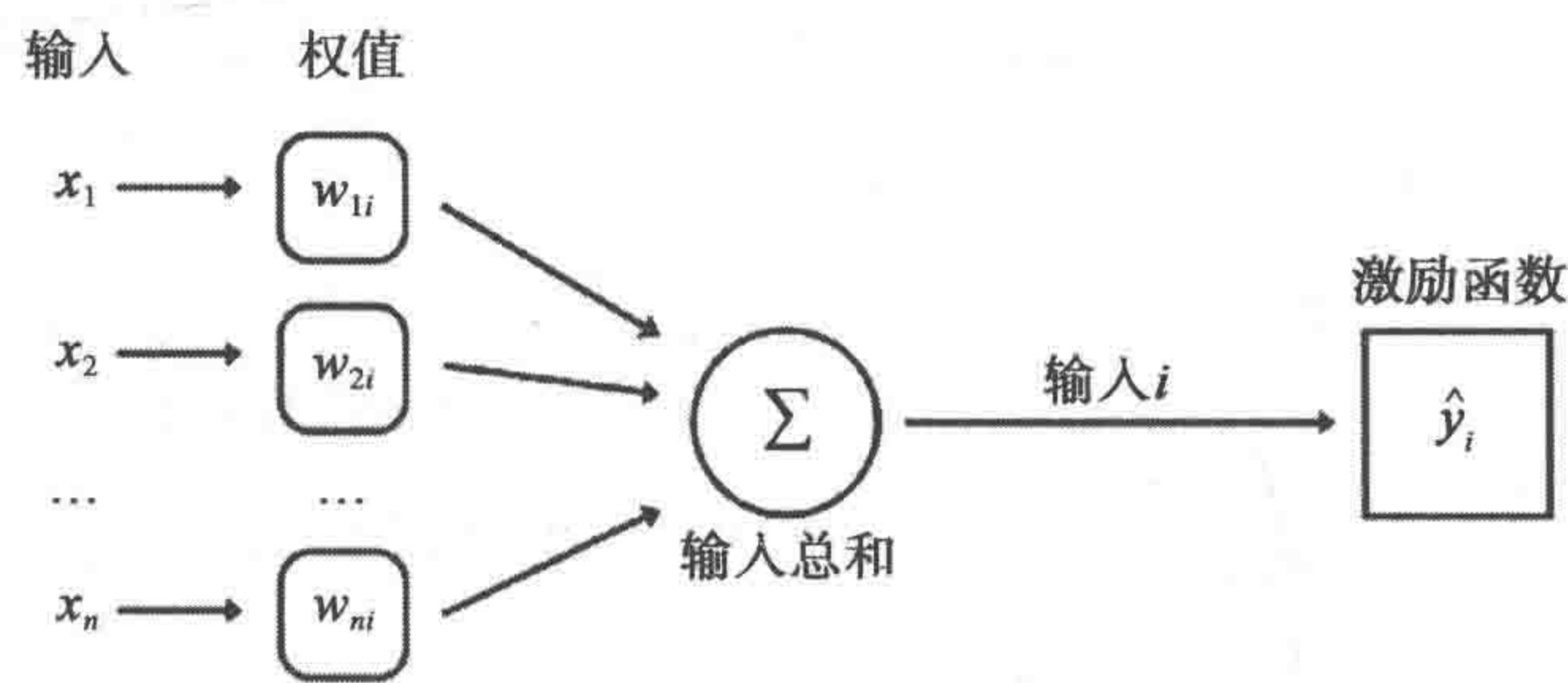


图 1-2 单层感知器模型

1.1.2 多层感知器模型

多层感知器（Multilayer Perceptron，MLP）模型和 SLP 模型非常相似，也是一种通过互联形成的前向反馈神经网络。每一层的神经元都与下一层神经元直接相连。MLP 和 SLP 之间最大的差别在于反向传播（Back-Propagation，BP）算法。BP 算法也是一种常见的神经网络训练方法，它将输出层的累积误差反向传递给输入层，以使用户能够了解每一层对整体误差的影响大小，并以此对网络进行调整。本节，我们将使用梯度下降算法来确定每次迭代中权值调整的大小。梯度下降（gradient descent）是另一种非常通用的机器学习/优化算法。梯度是函数的方向导数，它是一个标量值（标量意味着该值没有方向属性，只有大小），方向由函数冲量变化最大的那一点确定。沿着梯度不断减小的方向，我们就可以实现持续优化，从而最终得到一个全局优化的结果（参见图 1-3）。

1.1.3 卷积神经网络

卷积神经网络（Convolutional Neural Networks，CNN）是图像处理和计算机视觉中

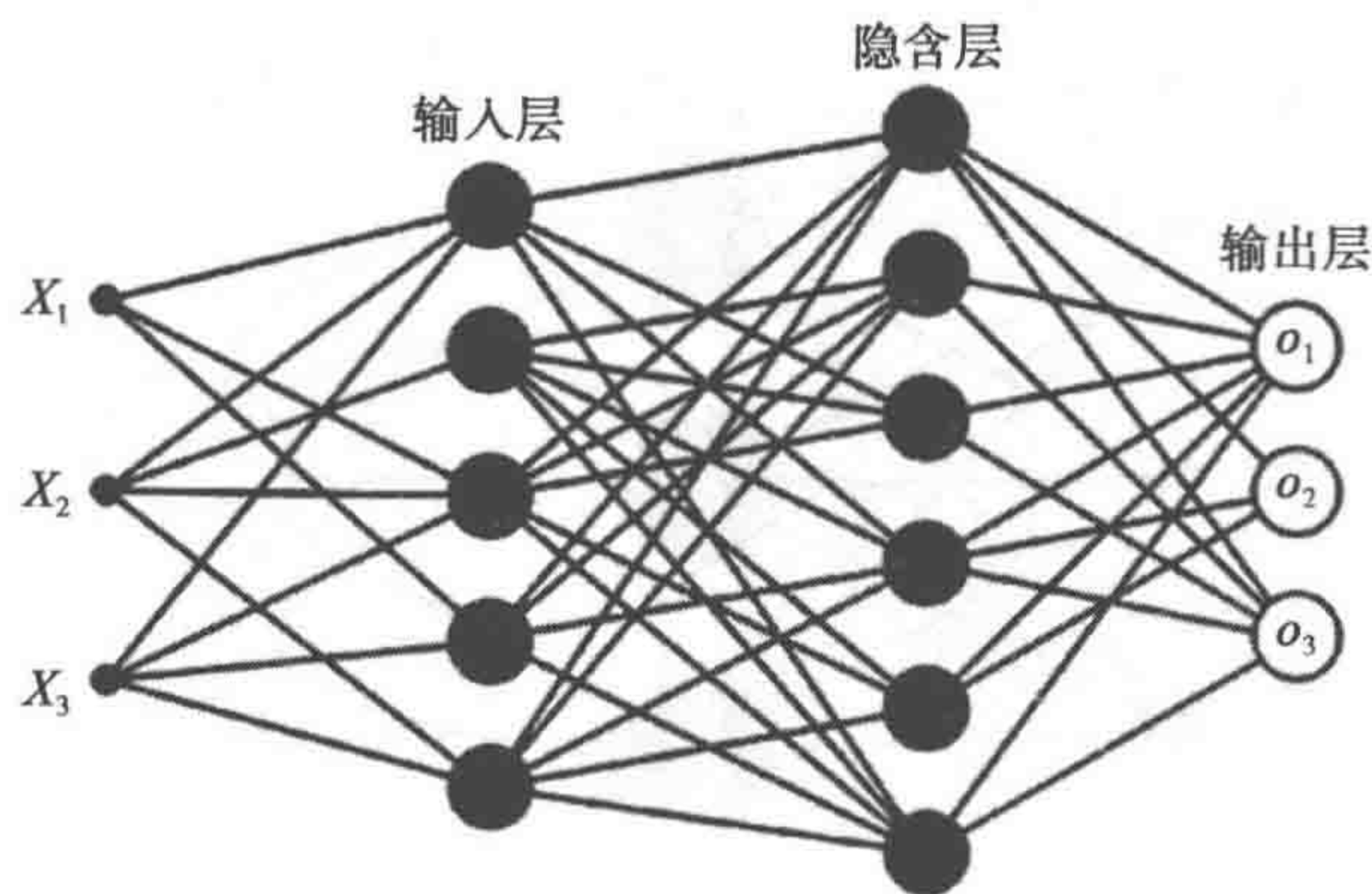


图 1-3 多层感知网络

最常用的模型。最初，研究人员希望通过 CNN 来模仿动物的大脑视觉皮层结构，所以不同于常规神经网络，CNN 的层具有按以下三维排列的神经元：宽度、高度和深度。每一层的神经元也仅仅与上一层中有限范围内的神经元相连接。CNN 模型最常用于图像处理和计算机视觉领域（参见图 1-4）。

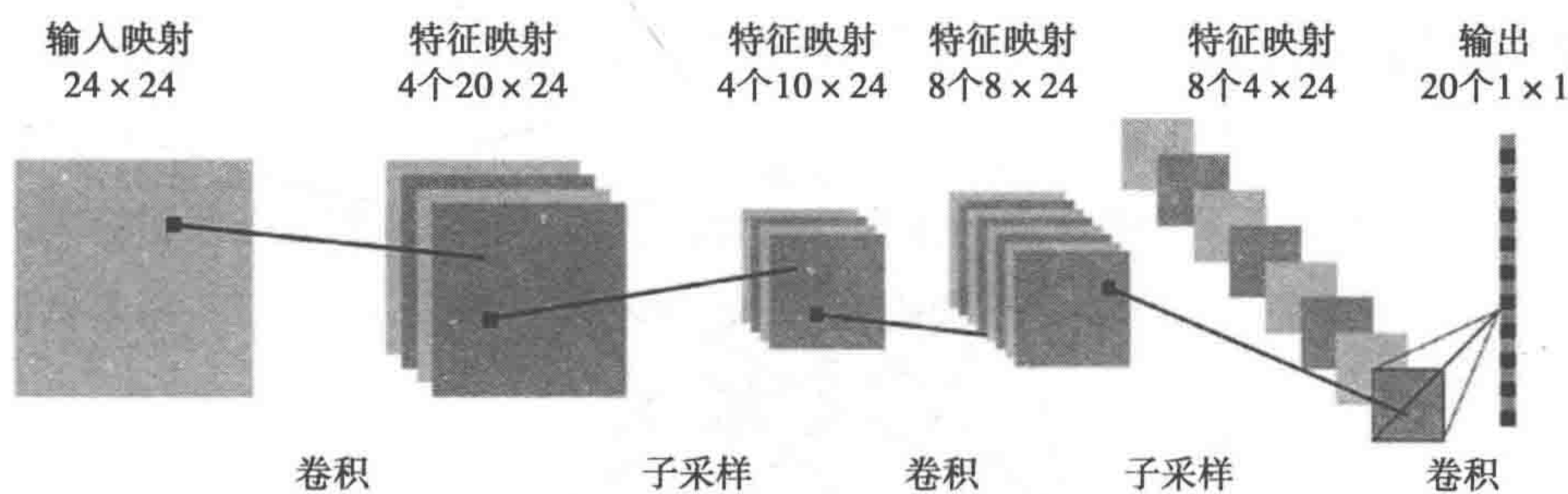


图 1-4 卷积神经网络

1.1.4 循环神经网络

循环神经网络（Recurrent Neural Networks, RNN）是一种包含定向环结构的人工神经网络（Artificial Neural Network, ANN）模型。其特别之处在于，RNN 中的定向环包含了网络的节点和边，连接顺序完全由边决定，因此表面上看来有时具备某个特定的顺序。RNN 较多应用在语音识别和手写识别这两个相关领域（参见图 1-5）。