



[PACKT]
PUBLISHING

深度学习系列



Deep Learning Essentials

深度学习

基础教程

邸 韩 (Wei Di)

[美]阿努拉格·巴德瓦杰 (Anurag Bhardwaj) 著

魏佳宁 (Jianing Wei)

杨 伟 李 征 等译



机械工业出版社
CHINA MACHINE PRESS



[PACKT]
PUBLISHING

深度学习系列

DEEP
LEARNING

Deep Learning Essentials

深度学习

基础教程

邸 韡 (Wei Di)

[美] 阿努拉格·巴德瓦杰 (Anurag Bhardwaj) 著

魏佳宁 (Jianing Wei)

杨 伟 李 征 等译



机械工业出版社
CHINA MACHINE PRESS

本书是真正适合深度学习初学者的入门书籍，全书没有任何复杂的数学推导。本书首先介绍了深度学习的优势和面临的挑战、深度学习采用深层架构的动机、学习深度学习需要的数学知识和硬件知识以及深度学习常用的软件框架。然后对多层感知机、卷积神经网络（CNN）、受限玻耳兹曼机（RBM）、循环神经网络（RNN）及其变体——长短时记忆（LSTM）网络进行了详细介绍，并且以独立章节重点阐述了CNN在计算机视觉中的应用、RNN在自然语言处理中的应用以及深度学习在多模态学习领域中的应用。最后，本书介绍了深度强化学习的基本知识，给出了应用深度学习技术需要的许多实用技巧并概述了深度学习的一些新方向和新应用。

Copyright © 2018 Packt Publishing

First published in the English language under the title “Deep Learning Essentials” / by Wei Di, Anurag Bhardwaj, Jianing Wei / ISBN: 978-1-78588-036-0

Copyright in the Chinese language (simplified characters) © 2018 China Machine Press

This translation of *Deep Learning Essentials* first published in 2018 is published by arrangement with Packt Publishing Ltd.

This title is published in China by China Machine Press with license from Packt Publishing Ltd. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由Packt Publishing Ltd授权机械工业出版社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）出版与发行。未经许可的出口，视为违反著作权法，将受法律制裁。

北京市版权局著作权合同登记 图字：01-2018-1733号。

图书在版编目（CIP）数据

深度学习基础教程 / (美) 邱鞞, (美) 阿努拉格·巴德瓦杰 (Anurag Bhardwaj), (美) 魏佳宁著; 杨伟等译. —北京: 机械工业出版社, 2018.10

(深度学习系列)

书名原文: Deep Learning Essentials

ISBN 978-7-111-60845-5

I. ①深… II. ①邱… ②阿… ③魏… ④杨… III. ①机器学习—教材 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2018) 第 208013 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑: 刘星宁

责任编辑: 刘星宁

责任校对: 王 延 陈 越 责任印制: 张 博

三河市宏达印刷有限公司印刷

2018 年 10 月第 1 版第 1 次印刷

184mm × 240mm · 11 印张 · 242 千字

0 001—4000 册

标准书号: ISBN 978-7-111-60845-5

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

电话服务

网络服务

服务咨询热线: 010-88361066 机工官网: www.cmpbook.com

读者购书热线: 010-68326294 机工官博: weibo.com/cmp1952

010-88379203 金书网: www.golden-book.com

封面无防伪标均为盗版

教育服务网: www.cmpedu.com

译者序 |

深度学习是采用多层神经网络学习数据层次表示的机器学习方法。尽管深度学习的基础理论突破在 20 世纪 90 年代之前就已完成，但由于当时计算机硬件和数据集大小的限制，深度学习相比于其他机器学习方法并没有展现出明显的优势。直到近年来随着海量带标签数据集的出现以及高性能 GPU 硬件的发展，深度学习才成为风靡全球的技术。目前，深度学习已经在人脸识别、图像检索、自动驾驶、机器翻译、语言识别等改变人类社会生活的领域获得了广泛的成功。为此，许多科技公司如谷歌、Facebook、微软、百度、腾讯和商汤科技等都在深度学习上投入了大量的资金和人力，以便抢占相关核心技术的制高点。

深度学习技术是构建在复杂的数学理论之上的，完全地掌握它需要线性代数、微积分、非线性优化以及概率论等方面的数学知识。这听起来往往令深度学习的初学者望而却步。然而，由于深度学习框架 TensorFlow、PyTorch、CNTK 和 Keras 等的出现，学习深度学习的门槛已经大大降低了。特别地，深度学习框架引入了自动微分技术，避免了用户基于微积分知识进行复杂的导数推导。另外，深度学习框架提供了多个成熟的优化算法，使得不了解非线性优化的用户可以通过简单的代码使用优化算法。

通过绕开复杂的数学推导，本书基于深度学习框架 TensorFlow 对深度学习的基本概念、常用网络模型和实践方法等进行了深入浅出的讲解。即便数学基础不好的读者，也可以轻松地阅读本书。本书首先介绍了多层感知机、卷积神经网络（CNN）、受限玻耳兹曼机（RBM）和循环神经网络（RNN），然后重点阐述了 CNN 如何被用于解决现实世界中的计算机视觉问题以及 RNN 及其变体——长短时记忆（LSTM）网络如何进行高级自然语言处理。最后，本书介绍了深度强化学习和深度学习的实用技巧，同时探讨了深度学习的发展趋势。特别地，通过设问及循序渐进的介绍，本书可以让读者深刻地领会当前的深度学习技术。不同于市面上的深度学习书籍，本书首次对深度学习在多模态领域中的应用进行了专门的介绍。

本书的作者 Wei Di、Anurag Bhardwaj 和 Jianing Wei 都是工业界的一线数据科学家。他们工作多年，具有深厚的理论功底和丰富的实践经验，这些都为本书的写作提供了良好的基础。

深度学习是人工智能领域的研究热点，随着时间的推移，各种新理论和新算法将不断涌现。本书只提供了迄今为止已经被证实的非常有效的深度学习技术。本书对深度强化学习的介绍略显单薄。尽管如此，本书仍然是不可多得的好教材。

本书由河南大学的杨伟、李征两位老师主译。在本书翻译过程中，河南大学计算机与信息工程学院的研究生李艳萍、韩森森、李聪颖、王玉、郭玥秀等也参加了部分内容的翻译并对译文的校订做了大量工作，机械工业出版社的刘星宁编辑在本书的整个翻译过程中

提供了许多帮助，在此对他们表示衷心的感谢。

此外，在翻译的过程中，我们对英文版中的少数排版错误进行了更正。由于错误非常明显，我们没有特别声明。译文虽经反复修改和校对，但由于译者水平有限，书中难免有欠妥和纰漏之处，我们真诚地欢迎广大读者批评指正。

译 者

原书前言 |

深度学习是科技界最具突破性的发展趋势，已经跨越研发实验室应用到生产环境中。它是通过数据的多个隐藏层进行深入洞察的科学和艺术。目前，深度学习是图像识别、物体识别和自然语言处理（NLP）问题解决方案的最佳提供者之一。

从温习机器学习开始，本书将快速进入深度学习的基本原理及其实现。随后，将向读者介绍不同类型的神经网络及其在现实世界中的应用。在具有启发性示例的帮助下，读者将学习使用深层神经网络识别模式，并了解数据操作和分类等重要概念。

采用基于深度学习的强化学习技术，将可以构建超越人类的人工智能（AI）。另外，将学习如何使用 LSTM 网络。阅读本书的过程中，将遇到各种不同的框架和库，如 TensorFlow、Python 和 Nvidia 等。在本书的最后，读者将能够为自己的应用部署一个面向生产的深度学习框架。

读者对象

如果你是一个希望为自己的商业应用构建深度学习动力的富有抱负的数据科学家、深度学习爱好者或人工智能研究人员，那么本书能够成为你开始处理人工智能挑战的完美资源。

为了充分利用本书，你必须具备中级 Python 技能，并且熟悉机器学习概念。

本书内容

第 1 章是为什么进行深度学习，对深度学习进行了概述。本章将首先介绍深度学习的历史、兴起及其在一些领域的最新进展。随后还将介绍深度学习的一些挑战及其未来潜力。

第 2 章是为深度学习做准备，是培养自己进行深度学习实验并在现实世界中应用深度学习技术的起点。本章将回答关于深度学习入门所需技能和概念的一些关键问题。本章内容包括线性代数的一些基本概念、深度学习实现的硬件要求，以及一些主流的深度学习软件框架。本章还将研究在基于云的 GPU 实例上从零开始搭建一个深度学习系统。

第 3 章是神经网络入门，重点介绍神经网络的基本知识，包括输入/输出层、隐藏层，以及网络如何通过前向和反向传播进行学习。本章将从标准多层感知神经网络及其构建模块开始，说明它们是如何逐步学习的。随后还将介绍一些主流的标准模型，如卷积神经网络（CNN）、受限玻耳兹曼机（RBM）、循环神经网络（RNN）及其变体——长短时记忆（LSTM）网络。

第 4 章是计算机视觉中的深度学习，对 CNN 进行了详细解释。本章将讲解 CNN 工作所必需的核心概念以及它们如何被用来解决现实世界中的计算机视觉问题。本章将介绍一些主流的 CNN 架构，并且还将使用 TensorFlow 实现一个基本的 CNN。

第 5 章是自然语言处理中的向量表示，涵盖了基于深度学习进行自然语言处理的基础

知识。本章将介绍自然语言处理中一些主流的用于特征表示的词嵌入技术，涵盖的模型有 Word2Vec、GloVe 和 FastText 等。本章还包括一个使用 TensorFlow 进行嵌入训练的示例。

第 6 章是高级自然语言处理，采用以模型为中心的方法进行文本处理。本章将讨论一些核心模型，如 RNN 和 LSTM 网络。本章将介绍使用 TensorFlow 实现一个 LSTM 网络的示例，并描述 LSTM 网络常用文本处理应用背后的基本架构。

第 7 章是多模态，介绍了采用深度学习处理多模态的一些基本进展。本章也分享了一些新颖、先进的深度学习多模态应用。

第 8 章是深度强化学习，涵盖了强化学习的基础知识，同时阐述了怎样应用深度学习改进强化学习。本章主要介绍了使用 TensorFlow 进行深度强化学习的基本实现，并讨论了深度强化学习的一些主流应用。

第 9 章是深度学习的技巧，为读者提供了在使用深度学习时可以采用的许多实用技巧，如网络权值初始化的最佳实践、学习参数调整、如何防止过拟合以及在面对数据挑战时如何准备数据以便更好地学习。

第 10 章是深度学习的发展趋势，总结了即将到来的一些深度学习想法。它着眼于新开发算法中的一些即将出现的趋势以及一些深度学习的新应用。

阅读建议

为了充分利用本书，读者需要按照我们提供的如下建议进行阅读：第一，建议至少对 Python 编程和机器学习有一些基本的了解。第二，在阅读第 3 章及其之后的章节之前，一定要遵循第 2 章的设置说明。需要设置自己的编程环境，以便于能够练习书中的示例代码。第三，让自己熟悉 TensorFlow 并阅读其文档。TensorFlow 文档 (https://www.tensorflow.org/api_docs/) 是一个很好的信息资源，其包含了大量很好且重要的示例。也可以在网上查看各种开源示例和深度学习相关的资源。第四，一定要亲自探索。对于不需要太多计算时间的简单问题尝试不同的设置或配置，这有助于快速了解模型如何工作以及如何调整参数。最后，深入研究每种类型的模型。本书用朴素的语言解释了各种深度学习模型的要点并且避免了太多的数学描述，其目的是帮助读者了解神经网络的底层机制。虽然目前有许多不同的开源工具可以提供高级 API，但是对深度学习的良好理解将对调试和改进模型性能大有裨益。

下载示例代码文件

读者可以在网站 www.packtpub.com 上登录自己的账户，下载本书的示例代码文件。如果是从其他渠道购买的本书，可以访问网址：www.packtpub.com/support。注册之后，可以收到通过邮件直接发送过来的文件。

可以根据下面的步骤下载代码文件：

- 1) 在网站 www.packtpub.com 上登录或注册；
- 2) 选择 **SUPPORT** 选项卡；
- 3) 单击 **Code Downloads & Errata**；

4) 在搜索框中输入本书的名字, 然后按照屏幕上的指令进行操作。

下载文件后, 请确保使用下列软件的最新版本解压或提取文件夹:

- Windows: WinRAR/7-Zip;
- Mac: Zippeg/iZip/UnRarX;
- Linux: 7-Zip/PeaZip。

本书的代码也可以通过 GitHub 网址: <https://github.com/PacktPublishing/Deep-Learning-Essentials> 进行下载。另外, 网址 <https://github.com/PacktPublishing/> 上含有大量其他书籍和视频中的代码, 欢迎下载!

下载彩色图片

作者还提供了一个 PDF 文件, 包含本书所用图片的彩色版本。读者可以通过访问网址: https://www.packtpub.com/sites/default/files/downloads/DeepLearningEssentials_ColorImages.pdf 进行下载。

排版约定

本书使用了许多文本约定。

正文中的代码、数据库表名、文件夹名、文件名、文件扩展名、路径名、虚拟 URL、用户输入和 Twitter 句柄都采用相同的字体排版。示例如下: “另外, `alpha` 是学习率; `vb` 是可见层的偏置; `hb` 是隐藏层的偏置; `W` 是权值矩阵。采样函数 `sample_prob` 是吉布斯采样函数, 其决定了要打开哪个节点。”

代码块的设置如下所示:

```
import mxnet as mx
tensor_cpu = mx.nd.zeros((100,), ctx=mx.cpu())
tensor_gpu = mx.nd.zeros((100,), ctx=mx.gpu(0))
```

任何命令行输入或输出具有如下形式:

```
$ sudo add-apt-repository ppa:graphics-drivers/ppa -y
$ sudo apt-get update
$ sudo apt-get install -y nvidia-375 nvidia-settings
```

新术语或重要词汇用黑体显示。



警告或者重要注释如左图所示。



提示和技巧如左图所示。

联系我们

我们欢迎来自读者的反馈。

一般反馈：请发送电子邮件至 feedback@packtpub.com，并在邮件中注明书名。如果读者对本书的任何方面有疑问，请发送电子邮件至 questions@packtpub.com。

勘误表：尽管我们已经尽全力确保内容的准确性，但错误不可避免。如果读者在本书中发现错误并告知我们，我们将不胜感激。请访问 www.packtpub.com/submit-errata，选择本书，单击勘误提交表单链接，然后输入详细信息。

盗版：如果读者在互联网上发现任何非法复制我们作品的情况，请将网址或网站名称提供给我们，我们将不胜感激。请通过 copyright@packtpub.com 与我们联系并提供材料链接。

成为作者：如果你有专业的知识，并且对撰写书籍感兴趣，请访问 authors.packtpub.com。

评论

请留下你的评论。一旦你阅读并使用了本书，为什么不在购买的网站上留下评论呢？潜在的读者可以看到并根据你中肯的意见做出购买决定，我们在 Packt 网站上可以了解你对我们产品的看法。同时，我们的作者可以看到你对他们著作的反馈。谢谢你！

有关 Packt 的更多信息，请访问 packtpub.com。

目 录

译者序

原书前言

第1章 为什么进行深度学习 // 1

1.1 什么是人工智能和深度学习 // 1

1.2 深度学习的历史及其兴起 // 2

1.3 为什么进行深度学习 // 7

1.3.1 相比于传统浅层方法的优势 // 7

1.3.2 深度学习的影响 // 8

1.4 深层架构的动机 // 9

1.4.1 神经的观点 // 10

1.4.2 特征表示的观点 // 10

1.5 应用 // 14

1.5.1 盈利性应用 // 14

1.5.2 成功案例 // 14

1.5.3 面向企业的深度学习 // 19

1.6 未来的潜力和挑战 // 20

1.7 小结 // 21

第2章 为深度学习做准备 // 22

2.1 线性代数的基础知识 // 22

2.1.1 数据表示 // 22

2.1.2 数据操作 // 23

2.1.3 矩阵属性 // 24

2.2 使用 GPU 进行深度学习 // 24

2.2.1 深度学习硬件指南 // 26

2.3 深度学习软件框架 // 27

2.3.1 TensorFlow // 27

2.3.2 Caffe // 27

2.3.3 MXNet // 28

2.3.4 Torch // 28

2.3.5 Theano // 28

2.3.6 CNTK // 29

2.3.7 Keras // 29

2.3.8 框架比较 // 29

2.4 基本亚马逊网络服务的深度学习开发环境配置 // 30

2.4.1 从零开始配置 // 30

2.4.2 基于 Docker 的配置 // 33

2.5 小结 // 34

第3章 神经网络入门 // 35

3.1 多层感知机 // 35

3.1.1 输入层 // 36

3.1.2 输出层 // 36

3.1.3 隐藏层 // 36

3.1.4 激活函数 // 36

3.2 如何进行网络学习 // 39

3.2.1 权值初始化 // 39

3.2.2 前向传播 // 39

3.2.3 反向传播 // 40

3.2.4 梯度消失与爆炸 // 42

3.2.5 优化算法 // 42

3.2.6 正则化 // 43

3.3 深度学习模型 // 43

3.3.1 卷积神经网络 // 43

3.3.2 受限玻耳兹曼机 // 47

3.3.3 循环神经网络 (RNN/LSTM 网络) // 50

3.3.4 RNN 中的单元及其展开 // 51

3.3.5 通过时间的反向传播 // 51

3.3.6 梯度消失与 LSTM 网络 // 51

3.3.7 LSTM 网络中的单元和网关 // 52

3.4 应用示例 // 53

3.4.1 TensorFlow 设置与关键概念 // 53

3.4.2 手写数字识别 // 54

3.5 小结 // 57

第 4 章 计算机视觉中的深度学习 // 58

4.1 卷积神经网络的起源 // 58

4.2 卷积神经网络 // 59

4.2.1 数据变换 // 61

4.2.2 网络层 // 62

4.2.3 网络初始化 // 65

4.2.4 正则化 // 66

4.2.5 损失函数 // 67

4.2.6 模型可视化 // 68

4.2.7 手写数字分类示例 // 69

4.3 微调卷积神经网络 // 71

4.4 主流的卷积神经网络架构 // 72

4.4.1 AlexNet // 72

4.4.2 VGG // 72

4.4.3 GoogLeNet // 73

4.4.4 ResNet // 73

4.5 小结 // 74

第 5 章 自然语言处理中的向量表示 // 75

5.1 传统的自然语言处理 // 75

5.1.1 BoW // 75

5.1.2 带权的 tf-idf // 76

5.2 基于深度学习的自然语言处理 // 77

5.2.1 动机及分散式表示法 // 77

5.2.2 词嵌入 // 78

5.2.3 Word2Vec // 81

5.2.4 了解 GloVe // 92

5.2.5 FastText // 92

5.3 应用 // 93

5.3.1 使用示例 // 93

5.3.2 微调 // 93

5.4 小结 // 94

第 6 章 高级自然语言处理 // 95

6.1 面向文本的深度学习 // 95

6.1.1 神经网络的局限性 // 95

6.2 循环神经网络 // 96

6.2.1 RNN 架构 // 97

6.2.2 基本的 RNN 模型 // 98

6.2.3 训练 RNN 很难 // 98

6.3 LSTM 网络 // 99

6.3.1 使用 TensorFlow 实现 LSTM 网络 // 101

6.4 应用 // 103

6.4.1 语言建模 // 103

6.4.2 序列标注 // 104

6.4.3 机器翻译 // 106

6.4.4 聊天机器人 // 109

6.5 小结 // 110

第 7 章 多模态 // 111

7.1 什么是多模态学习 // 111

7.2 多模态学习的挑战 // 111

7.2.1 模态表示 // 112

7.2.2 模态转换 // 112

7.2.3 模态对齐 // 112

7.2.4 模态融合 // 113

7.2.5 协同学习 // 113

7.3 图像标注生成 // 113

7.3.1 show&tell 算法 // 114

7.3.2 其他类型的方法 // 116

7.3.3 数据集 // 117

7.3.4 评估方法 // 119

7.3.5 注意力模型 // 121

7.4 视觉问答 // 125

7.5 基于多源的自动驾驶 // 127

7.6 小结 // 130

第 8 章 深度强化学习 // 131

8.1 什么是强化学习 // 131

8.1.1 问题设置 // 132

8.1.2 基于值函数学习的算法 // 132

8.1.3 基于策略搜索的算法 // 134

8.1.4 基于 Actor-Critic 的算法 // 134

8.2 深度强化学习 // 135

8.2.1 深度 Q 网络 (DQN) // 135

8.2.2 双 DQN // 137

8.2.3 竞争 DQN // 138

8.3 强化学习实现 // 138

8.3.1 简单的强化学习示例 // 138

8.3.2 以 Q-learning 为例的强化学习 // 139

8.4 小结 // 141

第 9 章 深度学习的技巧 // 142

9.1 处理数据 // 142

9.1.1 数据清理 // 142

9.1.2 数据扩充 // 142

9.1.3 数据归一化 // 143

9.2 训练技巧 // 143

9.2.1 权值初始化 // 143

9.2.2 优化 // 144

9.2.3 损失函数选择 // 146

9.2.4 防止过拟合 // 147

9.2.5 微调 // 148

9.3 模型压缩 // 149

9.4 小结 // 152

第 10 章 深度学习的发展趋势 // 153

10.1 深度学习的最新模型 // 153

10.1.1 生成对抗网络 // 153

10.1.2 Capsule 网络 // 154

10.2 深度学习的新应用 // 155

10.2.1 基因组学 // 155

10.2.2 医疗预测 // 157

10.2.3 临床影像学 // 159

10.2.4 唇语 // 160

10.2.5 视觉推理 // 161

10.2.6 代码合成 // 162

10.3 小结 // 164

第 1 章

为什么进行深度学习

本章给出了深度学习的简介，深度学习的历史及其兴起，以及深度学习在一些领域的最新进展。同时，讨论了深度学习的一些挑战及其未来潜力。

本章将回答一些关键的问题，这些问题通常由深度学习的实际用户提出，并且这些用户可能没有机器学习的背景。这些问题包括：

- 什么是人工智能（AI）和深度学习？
- 深度学习或人工智能的历史是什么？
 - 深度学习有什么重大突破？
 - 深度学习最近兴起的主要原因是什么？
- 深层网络结构的动机是什么？
 - 为什么要依靠深度学习，并且为什么现有的机器学习算法不能解决当前的问题？
 - 深度学习可以应用到哪些领域？
 - 深度学习的成功案例。
- 深度学习的未来如何及其当前的挑战是什么？

1.1 什么是人工智能和深度学习

创造模仿人类智能的梦想早已存在。虽然它们大多出现在科幻小说中，但在最近的几十年里，我们逐渐在建造智能机器上取得了进展，这些机器可以像人类一样完成某些任务。这就是称为人工智能的领域。人工智能的起源也许可以追溯到帕梅拉·麦考达克的著作《机器思维》（Machines Who Think），她在书中把人工智能描述成了能伪造神灵的古老愿望。

深度学习是人工智能的一个分支，其目标是让机器学习更接近于它最初的目标：人工智能。

深度学习采用的方法是模仿新皮层中神经元层中的活动。新皮层是占据大脑 80% 的褶皱，能够产生思维。在人类大脑中，大约有 1000 亿个神经元和 100~1000 万亿个突触。

对于不同类型的数据，如图像、视频、声音和文本，深度学习通过学习层次结构、表示级别和抽象级别理解数据模式。

高级抽象可定义为低级抽象的组合。前者之所以被称为深，是因为它有多个非线性特征变换状态。深度学习的最大优势之一就是它能够在多个抽象级别自动地学习特征表示。这使得系统不需要依赖于人工的特征提取，就能学习从输入空间到输出空间的复杂

函数映射。此外，深度学习还提供了预训练机制，即在一组数据集上学习的表示，可以应用到其他数据集。当然，预训练机制可能存在一些限制，比如用于学习的数据需要具备足够好的质量。此外，当以贪婪的方式基于大量无监督数据进行学习时，深度学习也会表现得很好。

图 1-1 显示了一个简化的卷积神经网络 (CNN)。

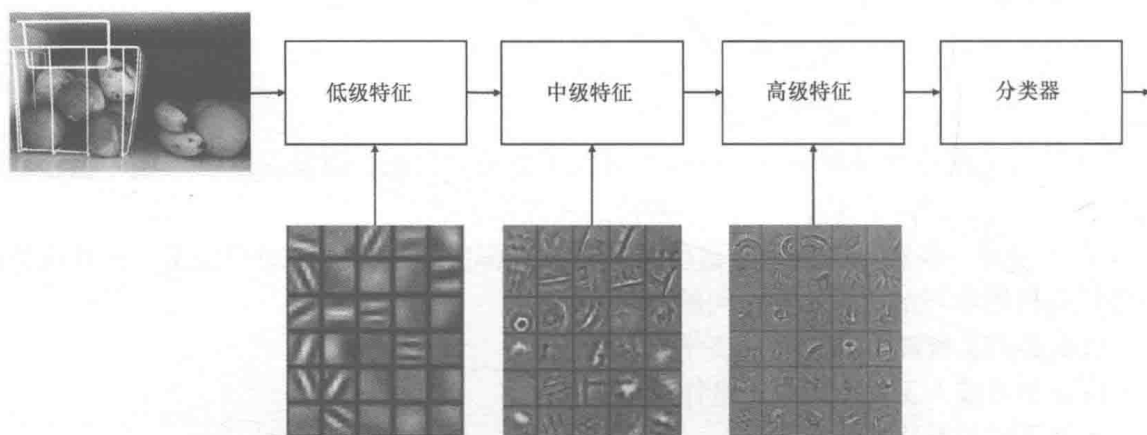


图 1-1 简化的卷积神经网络（每个小正方形中的图表示网络学到的层次特征）

深度学习模型，即学习后的深层神经网络，通常由多个网络层组成。这些网络层通过分层合作可以构建改进的特征空间。第一层学习低阶特征，例如颜色和边缘。第二层学习高阶特征，例如角点。第三层学习小块或纹理特征。网络层通常以无监督模式学习，以便发现输入空间的一般特征。然后，最后一层网络的特征可以输入到监督层以便完成分类或回归任务。

在网络层之间，节点通过加权边进行连接。每个节点与一个激活函数相关联，其可视为一个模拟的新皮层。节点的输入来自于其低层节点。然而，构建如此庞大、多层类神经元信息流阵列是 10 年前的想法。从创造想法到最近的成功，深层网络的发展经历了一波三折。

随着数学公式的最新改进、计算机的日益强大和越来越多的大规模数据集创建，深度学习的春天已经来临。目前，深度学习已成为当今科技界的一大支柱，并已广泛地应用于多个领域。在下一节中，将追溯深度学习的历史并讨论其令人难以置信的起伏发展旅程。

1.2 深度学习的历史及其兴起

在人工智能研究的曙光到来不久，20 世纪 40 年代便开发出了最早的神经网络。在 1943 年发表的题为《A Logical Calculus of Ideas Immanent in Nervous Activity》的开创性论文中，提出了神经网络的第一个数学模型。该模型的单元是一个简单的形式化神经元，通常被称为 McCulloch-Pitts 神经元。它是为生物神经元模型——神经网络而设想的数学函数。McCulloch-Pitts 神经元是人工神经网络中的基本单元。图 1-2 给出了一个人工神经

元模型的示例。人工神经元的想法确实很有发展前景，因为其以极其简单的方式模拟人脑的工作机制。

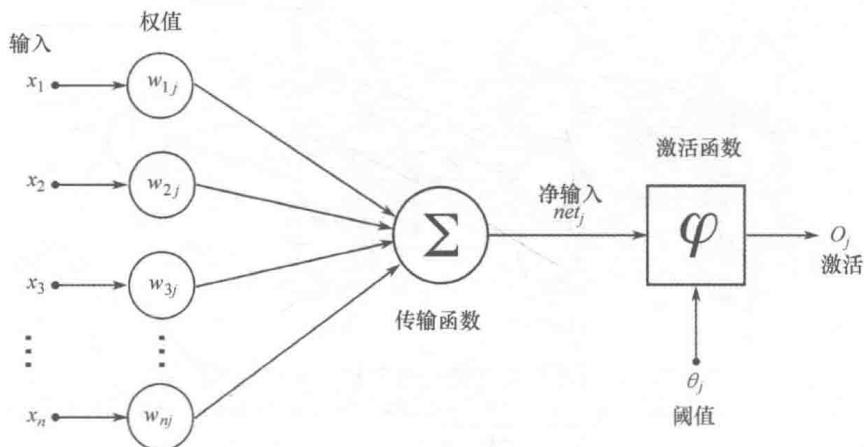


图 1-2 人工神经元模型示例

(图片来源: https://commons.wikimedia.org/wiki/File:ArtificialNeuronModel_english.png)

这些早期的模型只包括一小部分的人工神经元以及用于神经元连接的随机权值。权值决定人工神经元彼此间如何传递信息，即每个神经元如何以 0~1 之间的值进行响应。通过这种数学表示，神经元的输出可以获取图像的边缘及形状，或者音素中某一频率的特定能级。图 1-2 给出了人工神经元的数学公式表示，其中输入对应于树突；激活函数控制神经元是否在到达阈值时触发；输出对应于轴突。然而，早期的神经网络只能模拟非常有限的神经元数量，所以采用这种简单的架构仅能识别有限的模式。因此，这些模型在 20 世纪 70 年代被搁置。

反向传播概念是在 20 世纪 60 年代首次提出的，其表示采用误差训练深度学习模型。随后提出的是多项式激活函数模型。通过缓慢的人工处理，每一层基于统计选择的最佳特征被传送到下一层。不幸的是，第一个人工智能的冬天开始了，而且持续了近 10 年。在这个早期阶段，尽管模仿人脑的想法听起来十分奇妙，但是人工智能程序的实际功能却非常有限。即使最好的程序也只能处理一些简单问题。更别说当时拥有的计算能力非常有限，并且只有很小的数据集。严冬的出现主要是因为人们对人工智能的期望太高。当结果未能如意时，对人工智能研究的批评和撤资便随之而来。

慢慢地，反向传播算法在 20 世纪 70 年代有了显著的发展，但直到 1985 年才被应用于神经网络。在 20 世纪 80 年代中期，Hinton 等通过“深层模型”再次激发了人们对神经网络的兴趣。“深层模型”有两个以上的隐藏层，能更好地利用多层神经元。图 1-3 给出了一个多层感知机神经网络的示例。就是在这时候，Hinton 与其合作者证实了神经网络中的反向传播算法能够生成有趣的表示分布 (https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf)。1989 年，Yann LeCun(<http://yann.lecun.com/exdb/publis/pdf/lecun-89e.pdf>) 在贝尔实验室展示了反向传播算法的第一个实际应用。他把反向传播算法

引入到卷积神经网络（CNN）以便理解手写数字，其想法最终演变成了一个读取手写支票号码的系统。

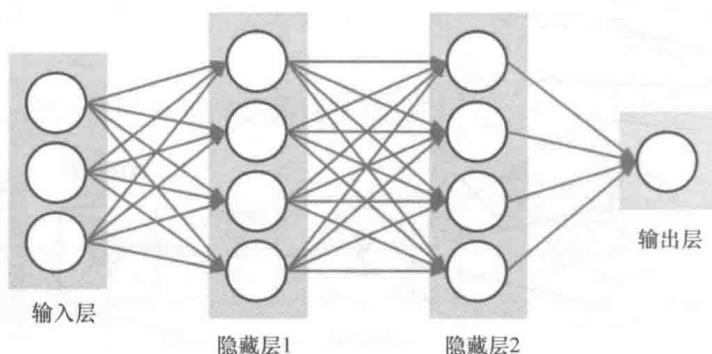


图 1-3 多层感知机神经网络中的人工神经元示例

（图片来源：https://github.com/cs231n/cs231n.github.io/blob/master/assets/nn1/neural_net2.jpeg）

此时恰逢第二个人工智能寒冬（1985~1990年）。1984年，两位主要的人工智能研究人员 Roger Schank 和 Marvin Minsky 对商界人士提醒到：人们对人工智能的热情已经走向失控。虽然多层网络可以学习复杂的任务，但是其运行速度很慢并且结果也不令人印象深刻。因此，当另一种较简单但更有效的方法（如支持向量机）被提出后，政府和风险投资者便放弃了对神经网络的支持。仅仅3年后，10亿美元的人工智能产业就崩溃了。

然而，这并不是真正的人工智能失败，而是更多炒作的结束。类似的情况在许多新兴技术中很常见。尽管人工智能在声誉、资金和兴趣方面起伏不定，一些研究人员仍然坚持自己的信念。不幸的是，他们并没有真正研究多层网络学习如此困难以及表现不佳的真实原因。2000年，梯度消失问题的发现最终引起人们关注到真正的关键问题：为什么多层网络不学习？其原因是一些激活函数的输入被压缩，即大面积的输入区域仅被映射到一个极小的输出区域。从最后一层网络计算出的较大变化或误差，只有少量会反向传播到前面的网络层。这意味着很少或根本没有学习信号能够到达前面的网络层，因而这些层学到的特征很弱。

需要注意的是，许多上层网络是问题的根本，因为它们承载了数据最基本的表征模式。由于上层网络的最优配置也可能依赖于其后续层的配置，因此这会使问题变得更糟。这意味着上层网络的优化是基于下层网络的非优配置进行的。所有这一切都表明要训练下层网络并获得好的结果是困难的。

针对上述问题，提出了两种解决方法：逐层预训练和长短时记忆（LSTM）模型。为了解决循环神经网络（RNN）的梯度消失问题，Sepp Hochreiter 和 Juergen Schmidhuber 在1997年首次提出了 LSTM 模型。

在最近的10年里，许多研究者取得了一些重要性的概念突破。无论学术界还是工业界，对深度学习的兴趣都突然爆发了。2006年，加拿大多伦多大学的 Hinton 教授与其合作者开发了一种针对深度置信网络（DBN）的快速学习算法（<https://www.cs.toronto.edu/~hinton/absps/fastnc.pdf>）。该算法能够更加有效地训练单个神经元层。这引发了神经网络的第二次复兴。在 Hinton 的论文中，其介绍了深度置信网络。深度置信网络在进行学习时一次贪婪

地训练一层，并且每层采用无监督学习算法——受限玻耳兹曼机（RBM）进行训练。图 1-4 阐明了这种深度置信网络的逐层训练概念。

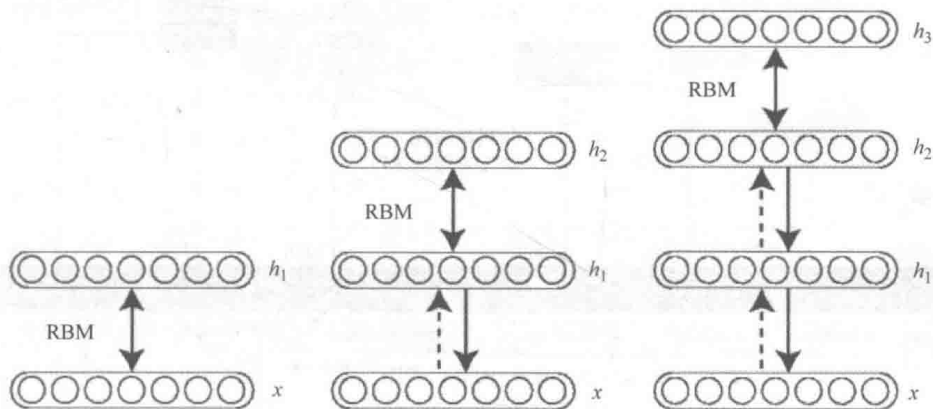


图 1-4 Hinton 提出的逐层预训练

提出的深度置信网络在 MNIST 数据集上进行了测试实验。MNIST 数据集是图像识别方法进行精确率和准确率比较的标准数据集，其由 70000 张大小为 28×28 像素的 0~9 之间的数字图片组成。70000 张图片被进一步划分为两部分：60000 张用于训练，余下的 10000 张用于测试。实验的目的是正确地回答测试样本中写入的数字是 0~9 之间的哪一个。虽然该论文在当时并没有引起太多关注，但是 RBM 在预测精确度方面远远高于传统的机器学习方法。

快进到 2012 年，一个算法震撼了整个人工智能研究领域。在全球图像识别竞赛——ImageNet 大规模视觉识别挑战（ImageNet Large Scale Visual Recognition Challenge, ILSVRC）中，一个名为 SuperVision 的冠军团队获得了 15.3% 的 top-5 测试错误率（<http://image-net.org/challenges/LSVRC/2012/supervision.pdf>），而第二名团队的测试错误率为 26.2%。ImageNet 数据集具有大约 120 万张高分辨率图片，涵盖了 1000 个不同的类别。ImageNet 提供了 1000 万张图片用作学习数据，15 万张图片用于测试数据。来自多伦多大学的 3 位作者 Alex Krizhevsky、Ilya Sutskever 和 Hinton 教授构建了一个具有 6000 万个参数、65 万个神经元以及 6.3 亿个连接的深层卷积网络。该网络由 7 个隐藏层和 5 个卷积层组成，其中部分卷积层后面紧跟着最大池化层。网络的最后是 3 个全连接层，最后一层全连接层的输出通过 1000 路的 softmax 函数最终产生 1000 个类别上的概率分布。通过随机地从输入图像中采样大小为 224×224 的图像块，实现了训练数据的扩充。为了加速训练，他们使用非饱和神经元和卷积运算的一个极高效的 GPU 实现。他们还在全连接层上使用 Dropout 以减少过拟合，这被证明是极其有效的。

此后，深度学习获得了飞速发展。现在，深度学习不仅成功地应用于图像分类，而且还在回归、维数约减、纹理建模、行为识别、运动建模、目标分割、信息检索、机器人、自然语言处理、语音识别、生物医学、音乐生成、艺术和协同过滤等领域获得了广泛应用。

图 1-5 给出了深度学习的发展路线图。有趣的是当我们回顾过去，似乎大多数的理论突破都已经在 20 世纪 80~90 年代完成。那么过去的 10 年里有什么改变呢？一个不太有争