



实战深度学习算法

零起点通关神经网络模型 (基于Python和NumPy实现)

徐彬◎著

实战深度学习算法

零起点通关神经网络模型 (基于Python和NumPy实现)

徐彬 著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

深度学习是机器学习的重要分支。本书系统地介绍了如何用 Python 和 NumPy 一步步地实现深度学习的基础模型，无须借助 TensorFlow、PyTorch 等深度学习框架，帮助读者更好地理解底层算法的脉络，进而进行模型的定制、优化和改进。全书由简到难地讲述感知机模型、多分类神经网络、深层全连接网络、卷积神经网络、批量规范化方法、循环神经网络、长短时记忆网络、双向结构的 BiGRU 模型等神经网络模型的必要算法推导、实现及其实例，读者可直接动手调试和观察整个训练过程，进一步理解模型及其算法原理。

本书适合没有深度学习基础，希望进入此领域的在校学生、研究者阅读，也适合有一定基础但不满足于“调包”和“调参”的工程师学习，还可供想要深入了解底层算法的研究人员参考阅读。

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

实战深度学习算法：零起点通关神经网络模型：基于 Python 和 NumPy 实现 / 徐彬著.

北京：电子工业出版社，2019.9

ISBN 987-7-121-37171-4

I. ①实… II. ①徐… III. ①机器学习—算法—研究 IV. ①TP181

中国版本图书馆 CIP 数据核字 (2019) 第 158449 号

责任编辑：孙学璜 sxy@phei.com.cn

印 刷：天津千鹤文化传播有限公司

装 订：天津千鹤文化传播有限公司

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：14 字数：280 千字 彩插：2

版 次：2019 年 9 月第 1 版

印 次：2019 年 9 月第 1 次印刷

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 zltz@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。

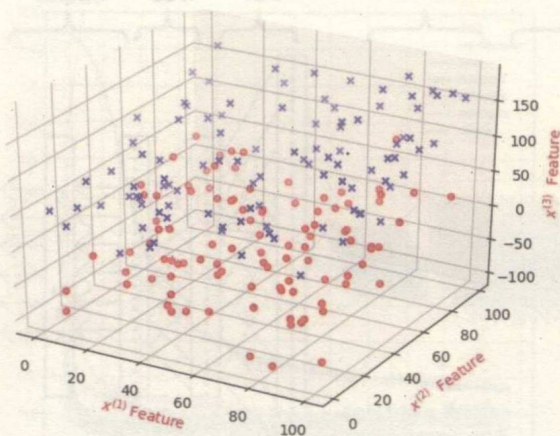


图1.1 三维空间中的线性可分样本

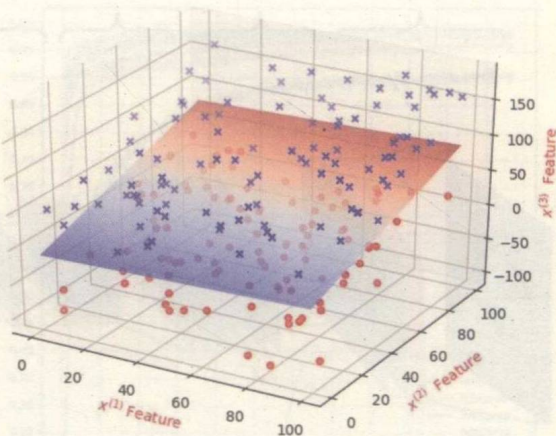


图1.4 空间中划分样本的一个分离平面

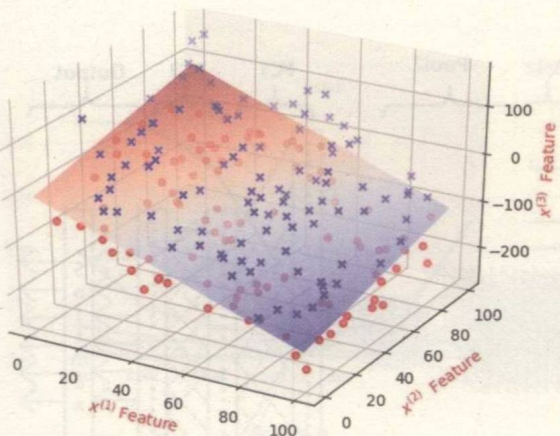


图3.1.1 空间旋转前，样本点在分离平面两侧

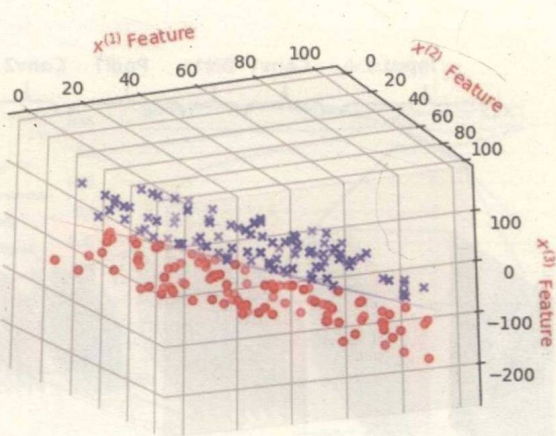


图3.1.2 空间旋转后，划分平面投影为一条直线，样本点的投影都在直线两侧

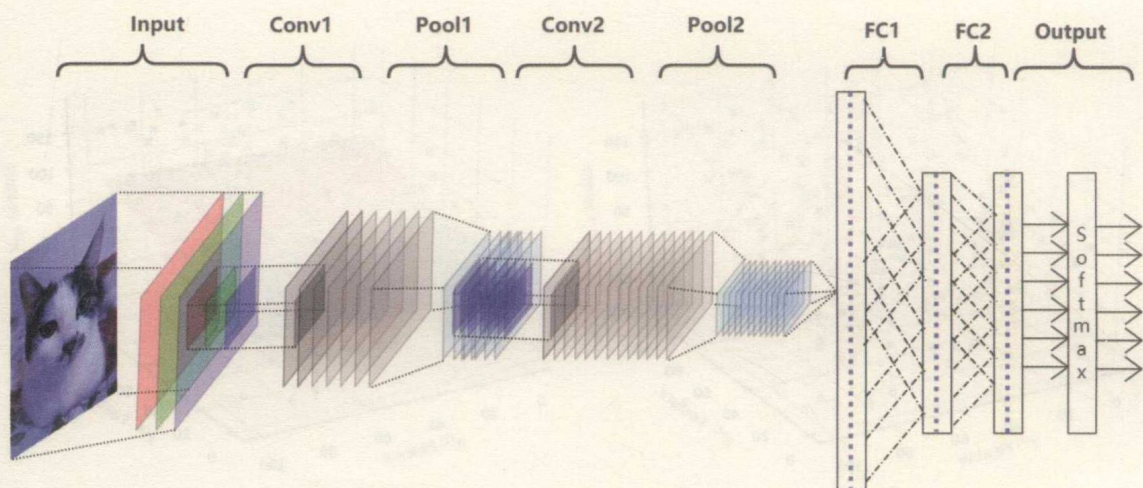


图5.5 卷积神经网络模型的结构，包含两组Conv-Pool单元

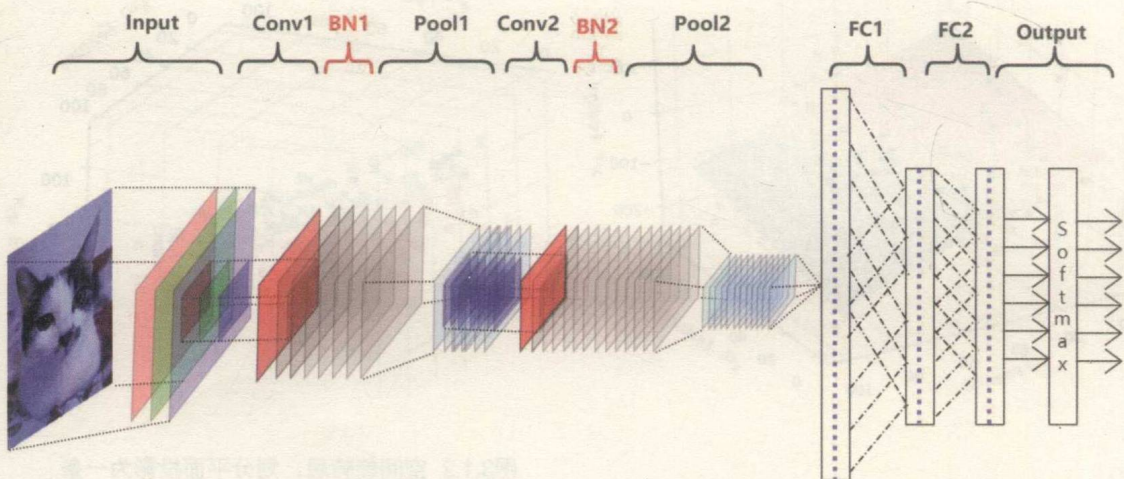


图6.4 卷积神经网络模型增加Batch Normalization层之后的结构

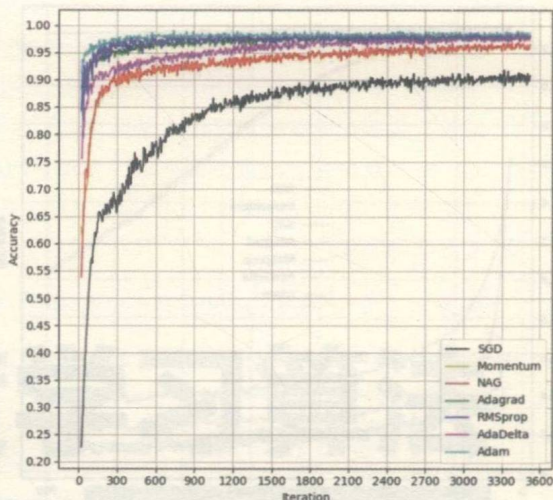
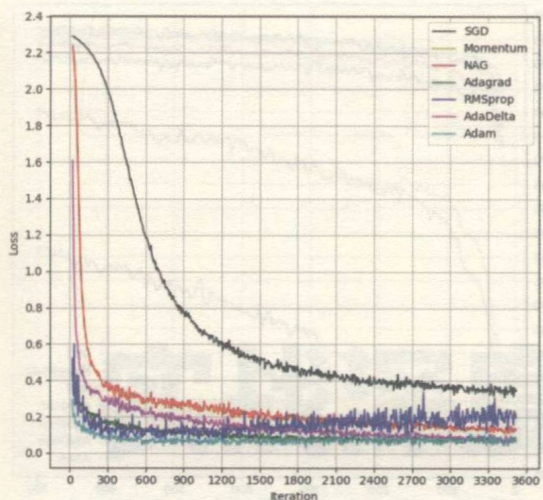
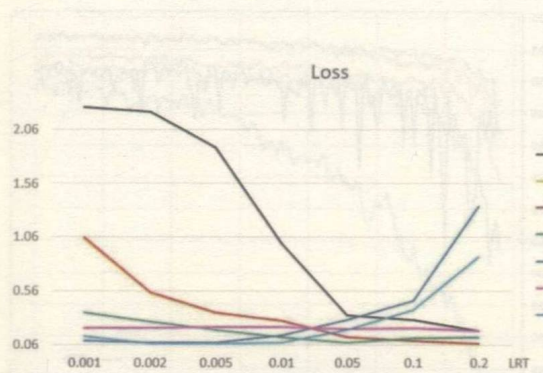
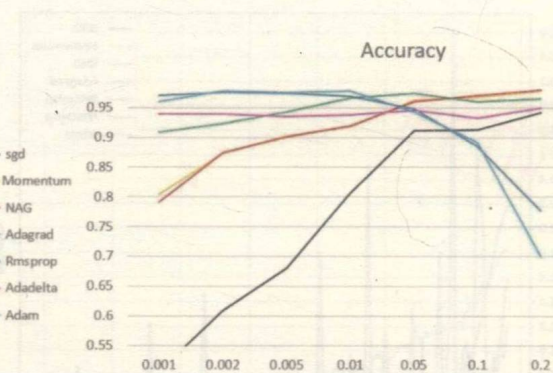


图5.1.1 合适的学习率下, 各种优化方法的损失

图5.1.2 合适的学习率下, 各种优化方法的正确率



左图5.3



右图5.4

不同的初始学习率下, Adam 方法和其他优化方法相比, 不见得更好, 反而可能是较差的

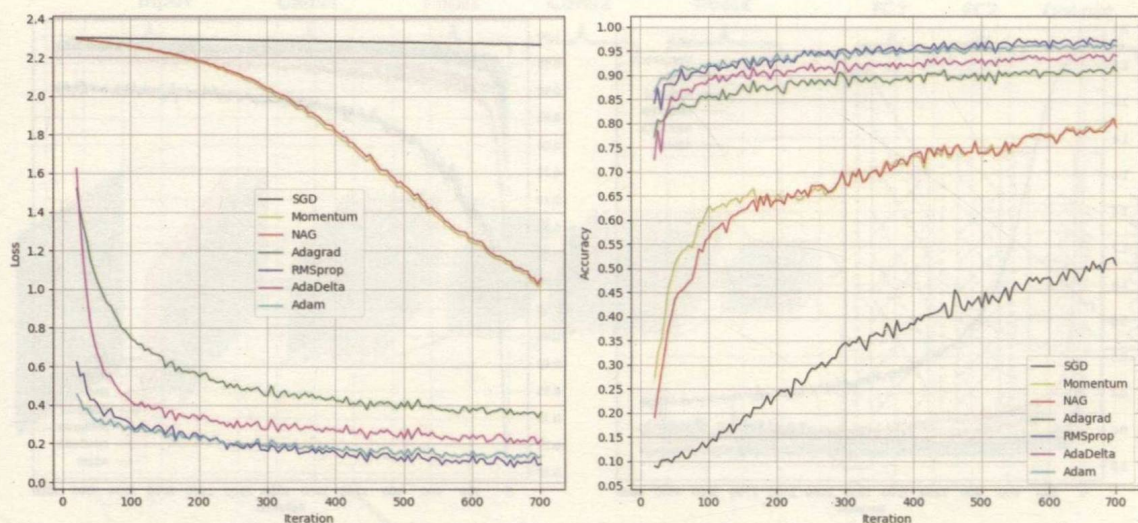


图5.7 各种优化方法在过小学习率下的表现

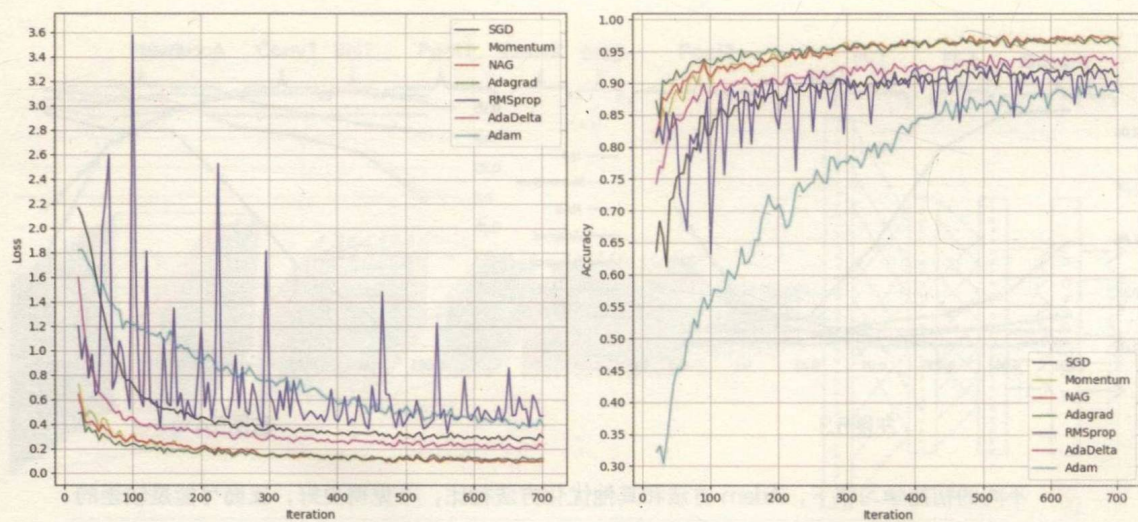


图5.8 各种优化方法在过大学习率下的表现

前言

以深度学习为代表的人工智能技术深刻地影响着我们的生活方式，从图像识别、语音识别、机器翻译到智能医诊、自动驾驶、智能风控……在多个应用领域不断地刷新纪录。

深度学习近年来之所以能取得颠覆性突破，一方面，归功于“数字化”对社会的渗透使得大量数据得以积累；另一方面，受益于单位成本下硬件算力的提升，推动了复杂模型的商用；然而最根本的，还是来自深度学习背后基础算法的巧思妙想与厚积薄发。

只有深入了解深度学习的算法原理，才能更灵活、高效地运用于实践当中。现有的深度学习框架将算法使用简化为“调包”和“调参”，降低了使用成本，然而却没有降低学习成本。对于算法，最有效的学习方式是理解原理并动手实践。从原始论文可以查阅算法的详解和推导，却不容易复现结果。主流的深度学习框架多采用计算图模型，不容易调试或观察，对希望深入了解算法的初学者并不友好。致力于用深度学习方法创造社会价值的从业者，也需要看清底层算法的脉络，来做模型的定制、优化和改进。

内容组织逻辑

本书的特点是“原理 + 实践”。按照“带着问题看算法”的逻辑来组织内容，所描述的每一种深度学习算法都围绕一个实际的目标问题展开，提供了基础算法的必要推导和实例，方便直观理解。

1. 提出问题。
2. 以问题为动机，引出模型。
3. 介绍模型原理、必要推导和实例。

4. 实现模型算法。
5. 解决问题与验证。

第 1 章至第 3 章，从感知机模型开始，步步渐进，介绍多分类神经网络、深层全连接神经网络；第 4 章至第 6 章，描述卷积神经网络（CNN）的核心算法、学习策略优化方法，以及深度学习的批量规范化方法；第 7 章至第 9 章，系统介绍了序列模型，基础的循环神经网络（Vanilla RNN）、长短时记忆网络（LSTM）和双向结构的 BiGRU 模型，以及序列模型适用的正则化方法。每章均以真实数据集作为目标问题，引出算法原理，不借助深度学习框架给予实现，最后完成数据集的验证。

阅读和使用

对于初学者，在阅读本书前，不需要具备机器学习基础，可以通过案例和模型概述等章节入门深度学习的概念；如果会使用 Python 语言简单编程，还可以结合书中的案例，动手了解各种模型所能解决的问题。

对于已有深度学习框架使用经验、希望了解底层算法的读者，可以查阅重要算法的原理、前向计算和反向传播的推导步骤，并在各章节的算法实现部分了解全部算法的实现过程。

附录部分包含了书中讨论的深度学习算法所涉及的数学基础，方便初学者速查和理解其直观意义。如需概念的严格定义和展开论证，可参考相关教材和专著。

通过阅读本书，希望读者可以：

- 理解深度学习主要的核心模型。
- 灵活复现重要论文、验证新方法。
- 自由替换模型中的底层算法，取得一手实验结果。
- 针对自己工作的特定场景，对算法做定制和优化，实现工程应用。

书中提供的算法实例基于 Python 和 NumPy 库实现，可以在 GitHub 公开代码库中获取 <https://github.com/Askylx/xDeepLearningBook>。

致谢

感谢为此书提供支持和做出过贡献的每个人。感谢家人的支持，他们是我动力的源泉。重庆大学郭平教授拨冗审阅了部分内容，提出了宝贵意见，对本书质量的提高有很大帮助，在此向他表示衷心的感谢。感谢一起工作过的同事和旅途巧遇的伙伴给予的启发，打开了前沿技术新世界的大门。电子工业出版社的孙学瑛女士在本书写作和出版过程中给予了很多帮助，在此特向她致谢。

由于作者水平有限，书中难免存在错误和不足，敬请专家和读者批评指正。如有意见或反馈，可以发邮件至 xu.bin.sh@foxmail.com。

也欢迎在知乎上交流 <https://www.zhihu.com/people/xu-jerry-82>。

徐彬

2019年7月

读者服务

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/37171>



符号表

a, x, w	标量
$\mathbf{a}, \mathbf{x}, \mathbf{w}$	n 阶张量 ($n \geq 1$)
$\vec{a}, \vec{x}, \vec{w}$	向量 (一阶张量)
M	矩阵 (二阶张量)
\mathbf{x}^T, M^T	向量 \mathbf{x} 和矩阵 M 的转置
$f : \rightarrow \mathbb{R}$	输出实数标量值的函数
$\mathbf{f} : \rightarrow \mathbb{R}^m$	输出 m 维实数向量的函数
$\text{diag}(f)$	函数 f 的值构成的对角方阵
I	单位阵
$f'(x)$	函数 f 的一阶导函数
\cdot	点积运算符
\oplus	矩阵对应位置元素相加运算符
\otimes	矩阵对应位置元素相乘运算符, 用于 Hadamard 乘积运算
\odot	互相关运算符, 用于神经网络中的卷积运算
\circledast	经典卷积运算符

符号表

$z^{(l)}$	神经网络第 l 层原始输出
$a^{(l)}$	神经网络第 l 层激活后输出
$\delta^{(l)}$	回传至神经网络第 l 层误差项

目录

第 1 章 基础分类模型	1
1.1 深度学习简介	2
1.2 目标问题：空间中的二分类	2
1.3 感知机模型	3
1.3.1 感知机函数	3
1.3.2 损失函数	4
1.3.3 感知机学习算法	6
1.4 算法实现	8
1.4.1 环境搭建	8
1.4.2 数据准备	9
1.4.3 实现感知机算法	11
1.5 小结	13
参考文献	13
第 2 章 第一个神经网络	14
2.1 目标问题：MNIST 手写数字识别	15
2.1.1 数据集	15
2.1.2 图像数据和图向量	16
2.2 挑战：从二分类到多分类	16
2.3 Softmax 方法	19
2.4 正确分类的独热编码	20
2.5 损失函数——交叉熵	21
2.6 信息熵和交叉熵	21

目录

2.6.1	信息熵	21
2.6.2	交叉熵	22
2.7	第一个神经网络的学习算法	23
2.8	反向传播	26
2.9	抽象泄漏	27
2.10	算法实现	28
2.10.1	数据准备	28
2.10.2	实现第一个神经网络	33
2.10.3	实现 MINIST 手写数字识别	36
2.11	小结	37
	参考文献	38
第 3 章	多层全连接神经网络	39
3.1	第一个挑战：异或问题	40
3.2	更深的神经网络——隐藏层	40
3.3	第二个挑战：参数拟合的两面性	42
3.4	过拟合与正则化	44
3.4.1	欠拟合与过拟合	44
3.4.2	正则化	44
3.4.3	正则化的效果	44
3.5	第三个挑战：非线性可分问题	45
3.6	激活函数	45
3.7	算法和结构	47
3.8	算法实现	50
3.8.1	数据准备	50
3.8.2	实现多层全连接神经网络	50
3.8.3	在数据集上验证模型	53
3.9	小结	54
	参考文献	54
第 4 章	卷积神经网络 (CNN)	55
4.1	挑战：参数量和训练成本	56
4.2	卷积神经网络的结构	56

4.2.1	卷积层	57
4.2.2	池化层	62
4.2.3	全连接层和 Softmax 处理	63
4.3	卷积神经网络学习算法	63
4.3.1	全连接层	63
4.3.2	池化层反向传播	64
4.3.3	卷积层反向传播	65
4.4	算法实现	68
4.4.1	数据准备	68
4.4.2	卷积神经网络模型的原始实现	69
4.5	小结	76
	参考文献	78
第 5 章	卷积神经网络——算法提速和优化	79
5.1	第一个挑战：卷积神经网络的运算效率	80
5.2	提速改进	80
5.2.1	边缘填充提速	82
5.2.2	池化层提速	83
5.2.3	卷积层处理	85
5.3	反向传播算法实现	88
5.3.1	池化层反向传播	88
5.3.2	卷积层反向传播	89
5.4	第二个挑战：梯度下降的幅度和方向	91
5.5	递减学习率参数	92
5.6	学习策略的优化方法	92
5.6.1	动量方法	93
5.6.2	NAG 方法	93
5.6.3	Adagrad 方法	94
5.6.4	RMSprop 方法	95
5.6.5	AdaDelta 方法	96
5.6.6	Adam 方法	97
5.6.7	各种优化方法的比较	98

目录

5.7	总体模型结构	100
5.8	使用 CNN 实现 MNIST 手写数字识别验证	101
5.9	小结	102
	参考文献	103
第 6 章	批量规范化 (Batch Normalization)	104
6.1	挑战: 深度神经网络不易训练	105
6.2	批量规范化方法的初衷	105
6.2.1	数据集偏移	106
6.2.2	输入分布偏移	106
6.2.3	内部偏移	107
6.3	批量规范化的算法	107
6.3.1	训练时的前向计算	107
6.3.2	规范化与标准化变量	108
6.3.3	推理预测时的前向计算	109
6.3.4	全连接层和卷积层的批量规范化处理	110
6.4	批量规范化的效果	111
6.4.1	梯度传递问题	111
6.4.2	饱和非线性激活问题	112
6.4.3	正则化效果	113
6.5	批量规范化为何有效	113
6.6	批量规范化的反向传播算法	114
6.7	算法实现	115
6.7.1	训练时的前向传播	116
6.7.2	反向传播	117
6.7.3	推理预测	118
6.8	调整学习率和总体结构	119
6.8.1	模型结构	119
6.8.2	卷积层批量规范化的实现	120
6.8.3	引入批量规范化后的递减学习率	121
6.9	在 MNIST 数据集上验证结果	122
6.10	小结	123

参考文献	123
第 7 章 循环神经网络 (Vanilla RNN)	125
7.1 第一个挑战: 序列特征的捕捉	126
7.2 神经网络的结构	126
7.2.1 单层 RNN	126
7.2.2 双向 RNN	128
7.2.3 多层 RNN	129
7.3 RNN 前向传播算法	130
7.4 RNN 反向传播算法	131
7.4.1 误差的反向传播	131
7.4.2 激活函数的导函数和参数梯度	132
7.5 第二个挑战: 循环神经网络的梯度传递问题	133
7.6 梯度裁剪	134
7.7 算法实现	135
7.8 目标问题: 序列数据分析	139
7.8.1 数据准备	139
7.8.2 模型搭建	144
7.8.3 验证结果	145
7.9 小结	147
参考文献	147
第 8 章 长短时记忆网络 (LSTM) ——指数分析	149
8.1 目标问题: 投资市场的指数分析	150
8.2 挑战: 梯度弥散问题	150
8.3 长短时记忆网络的结构	150
8.4 LSTM 前向传播算法	152
8.5 LSTM 反向传播算法	153
8.5.1 误差反向传播	154
8.5.2 激活函数的导函数和参数梯度	155
8.6 算法实现	156
8.6.1 实现 LSTM 单时间步的前向计算	156
8.6.2 实现 LSTM 多层多时间步的前向计算	157