

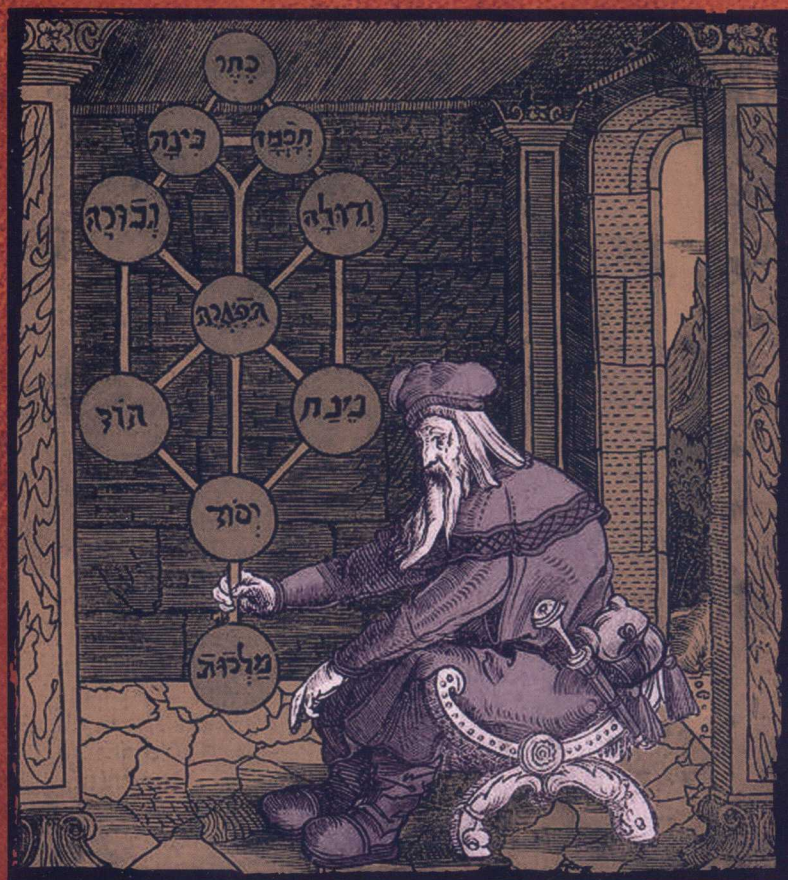
TURING

图灵原创

深入理解神经网络

从逻辑回归到CNN

张觉非 著



中国工信出版集团

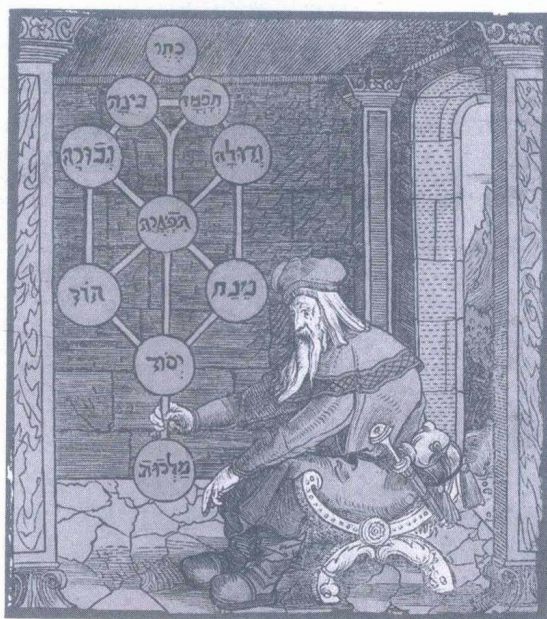


人民邮电出版社
POSTS & TELECOM PRESS

深入理解神经网络

从逻辑回归到CNN

张觉非 ◎ 著



人民邮电出版社
北 京

图书在版编目(CIP)数据

深入理解神经网络：从逻辑回归到CNN / 张觉非著.
— 北京：人民邮电出版社，2019.9
(图灵原创)
ISBN 978-7-115-51723-4

I. ①深… II. ①张… III. ①人工神经网络—研究
IV. ①TP183

中国版本图书馆CIP数据核字(2019)第155659号

内 容 提 要

本书以神经网络为线索，沿着从线性模型到深度学习的路线讲解神经网络的原理和实现。本书将数学基础知识与机器学习和神经网络紧密结合，包含线性模型的结构与局限、损失函数、基于一阶和二阶信息的优化算法、模型自由度与正则化、神经网络的表达能力、反向传播与计算图自动求导、卷积神经网络等主题，帮助读者建立基于数学原理的较深刻的洞见和认知。本书还提供了逻辑回归、多层全连接神经网络和多种训练算法的Python实现，以及运用TensorFlow搭建和训练多种卷积神经网络的代码实例。

本书适合渴望加深对神经网络和深度学习原理理解的高年级本科生与研究生、广大程序员与工程师，以及对机器学习的原理和编程实现感兴趣的所有读者阅读。

-
- ◆ 著 张觉非
责任编辑 陈兴璐
责任印制 周昇亮
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京市艺辉印刷有限公司印刷
 - ◆ 开本：800×1000 1/16
印张：20.25
字数：479千字
印数：1—3 000册
- 2019年9月第1版
2019年9月北京第1次印刷
-

定价：89.00元

读者服务热线：(010)51095183转600 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字 20170147 号



— 张觉非 —

本科毕业于复旦大学计算机系，
于中国科学院古脊椎动物与
古人类研究所取得古生物学硕士学位，
目前在互联网行业
从事机器学习算法相关工作。

站在巨人的肩上
Standing on Shoulders of Giants



iTuring.cn

献给周怡萍女士，我生命中的奇异吸引子。

前言

“理论是灰色的，我亲爱的朋友，而生命的金树长青。”

——歌德，《浮士德》

近年来，深度学习的浪潮席卷了学术界和产业界，它强大的能力给人们留下了深刻的印象。在产业界，一些已经成熟应用机器学习技术的领域，如计算广告、推荐系统、机器视觉、自然语言处理等，纷纷尝试深度学习并取得了良好的效果。各个尚未应用机器学习的领域也都在积极进行智能化的尝试。另外，开源社区贡献的众多框架、工具和平台，也为这股浪潮推波助澜。可以说，机器学习，尤其是神经网络与深度学习，是当今最被寄予厚望的技术。

不论这股浪潮将止于何处，也不论它最终能否达到人们的期许，机器学习的思想都已经深深植入了人们的心中。在工程和业务实践中，将数据和建模纳入解决方案已是一种常规做法。机器学习已成为工程师武器库中一个重要的工具，越来越多工程师开始尝试了解和应用机器学习技术。

但是人们会发现，机器学习有着艰深的理论背景，不能透彻理解原理，就难以在实践中自由地运用。当人们试图进入机器学习领域时，数学原理是横亘在面前的一座大山。笔者通过自身经历以及与他人的交流，深切体会到，广大学生和工程师在基础理论知识的理解和运用上普遍有所欠缺，这给他们理解机器学习的原理制造了障碍。

举个例子，对于梯度和链式法则，许多人只能在一元函数情况下理解，大量文章和书籍也都只在一元情况下做浅尝辄止的讲解。而当人们阅读代码时，会发现其中满是矩阵和向量。原理与实现之间仿佛弥漫着一团迷雾，不穿透这团迷雾，对这些概念的理解就只能是浮光掠影、隔靴搔痒。笔者曾读到有文章将梯度下降的矩阵实现称为“向量化”。其实何须向量化，多元函数的梯度与求导原本就是用矩阵和向量语言来描述的。

再比如，神经网络与深度学习的火热虽然给机器学习这门学科带来进步和活力，但是也给人们造成一个错误的印象，即它是全新的、颠覆性的、与传统割裂的。其实神经网络与深度学习深

深扎根于传统机器学习之中。读者可以看到，本书的大量篇幅都在讲解线性模型，这是因为神经网络与深度学习的结构和训练方法都扎根于线性模型之中。

目标读者

笔者遇到过不少理工科相关专业的工程师，他们学习过微积分、线性代数和概率论等基础课程，但是对这些知识的理解尚没有融会贯通，不知如何运用，而且搁置多年，遗忘严重。还有一些高年级本科生和研究生感觉各门基础课程仿佛是分散的、孤立的，不理解它们的联系和用途。本书的目标就是帮助这些读者回忆并夯实基础知识，深入理解机器学习，特别是神经网络和深度学习的原理。

本书以神经网络为线索，沿着从线性模型到深度学习的路线，串起核心知识点。数学内容在必要和恰当的时机引入，从基础讲起，不留黑盒。具有理工科背景的读者能够容易地回忆起相关知识，而不需要再回头求诸于大部头教科书。本书内容取舍有当，紧紧围绕主题，使读者能将数学与它在机器学习中的应用紧密结合起来。

本书试图帮助读者打通关节，提高视角，加深洞见，做到知其然并知其所以然。在当今人工智能时代，新方法不断涌现，新工具层出不穷，但是万变不离其宗，新的方法和工具都深深扎根于原理之中。把众多方法比作岛屿，在水面之下看，大大小小的岛屿就不再是孤立的、个别的，它们都建筑于基岩之上，基岩就是机器学习的基础理论。掌握了基础理论，就可以对各种方法形成统一而深刻的洞见，在工程实践中拥有坚实有效的理论依据，并能够快速理解和运用业界新涌现的众多方法和工具。

具体来说，本书适合以下几类读者：

- 渴望进入这一领域的高年级本科生和研究生，本书帮助这类读者将数学基础知识与机器学习和神经网络结合起来，深入理解原理；
- 希望了解神经网络与机器学习的广大程序员与工程师，这类读者需要花更多一些力气回忆数学知识，本书包含了他们需要的全部知识点；
- 对模型的编程实现感兴趣的读者，本书包含逻辑回归、多层全连接神经网络以及各种训练算法的 Python 实现，可供这类读者学习和参考；
- 业界的机器学习、数据挖掘工程师，本书关于模型原理的高级主题，特别是关于模型自由度与偏置-方差权衡方面的内容，可为他们提供一些有趣的洞见。

明确本书不包含的内容以及不适合的读者同样重要，以下是本书不涉及的内容：

- 本书围绕着神经网络这个主题，并涵盖了机器学习学科的相当一部分重点内容，但并不涵盖机器学习的全部领域；
- 本书包括逻辑回归、多层全连接神经网络以及多种训练算法的 Python 实现，但书中代码实现的目的是为了理解原理，本书并不是一本实现各类机器学习模型的指南；
- 本书包含使用 TensorFlow 搭建并训练模型的实例，但本书远远不是一本 TensorFlow 的实用教材；
- 本书从原理上讲解了模型超参数、自由度、过拟合与欠拟合等概念，但本书并非关于调参和优化模型的实践指南，也不包含特征预处理、特征工程等方面的内容；
- 本书以卷积神经网络为例讲解深度学习，涵盖了训练深度神经网络的方法、问题以及技术，但本书并不包含深度学习众多五花八门的新应用领域和网络结构。

内容概览

本书沿着从线性模型到神经网络，再到深度学习的路径，层层递进，逐渐深入。这是一条从根至叶、自简入繁的天然路径。各章节的安排如下。

- 第1章以逻辑回归为例讲解线性模型。这一章在引入逻辑回归的定义后回顾必要的向量几何，使读者对逻辑回归的原理、特性和局限形成清晰的认识。
- 第2章介绍模型训练和评价的基本概念，并引入损失函数。损失函数将模型训练问题转化为函数优化问题。这一章以交叉熵为例讲解损失函数，并从信息论、贝叶斯和几何特性三个角度探究交叉熵的原理。
- 第3章介绍基于函数局部一阶信息的优化算法。这一章首先回顾多元函数微分的相关知识并引入梯度概念，接着讲解梯度下降法及其变体，最后介绍如何用梯度下降法训练逻辑回归模型。
- 第4章介绍基于函数局部二阶信息的优化算法。这一章首先回顾矩阵的相关知识，接着讲解多元函数的赫森矩阵，以牛顿法和共轭方向法为例介绍二阶优化算法，最后介绍如何用牛顿法训练逻辑回归模型。
- 第5章首先回顾概率论相关知识，之后从线性回归入手，介绍模型自由度、正则化和偏置-方差权衡等概念，最后介绍如何在逻辑回归的训练中应用正则化。
- 第6章讲解如何将线性模型连接成网络以克服其局限，并介绍多层全连接神经网络。
- 第7章讲解训练多层全连接神经网络的反向传播加梯度下降法。这一章是对之前各章节知识的一个综合，最后还谈到了训练深层网络所面临的一些困难。

- 第8章介绍计算图和自动求导。深度学习中的各种网络具有比多层全连接神经网络复杂得多的连接方式，而计算图和自动求导是搭建和训练复杂网络结构的有力工具。
- 第9章介绍卷积神经网络的原理、结构和训练。深度神经网络五花八门，卷积神经网络是其中最典型、最具代表性的一种。理解了卷积神经网络及其训练，也就能够理解其他各种深度神经网络。
- 第10章介绍了5种经典的卷积神经网络，介绍它们的结构和性能，最后简单讨论卷积神经网络结构的发展演化趋势。
- 第11章展示如何使用TensorFlow搭建和训练本书提到的几种主要模型，并将它们用于手写数字识别。
- 在附录中，我们尝试在卷积神经网络和元胞自动机之间建立联系，从动力学角度理解“深度”的含义。元胞自动机是一类计算模型，它们中的一些具备图灵完备性。元胞自动机的计算方式与卷积具有相似性，这一章从元胞自动机的动力学特性的视角看卷积神经网络，希望为读者提供一些有趣的洞见。

阅读方式

本书各章之间存在依赖关系，但对于不同背景 and 需求的读者，可以选择不同的阅读路径。

- 数学基础过硬的读者可以略过1.2节、3.1节、4.1节、5.1节和7.1节，但温习一下总是有益的。
- 熟悉传统机器学习，想进一步了解神经网络和深度学习的读者，可以略过第一部分。
- 对于第4章，读者可以略过4.1节之后的内容，因为后续章节没有用到二阶优化算法。但4.1节回顾的矩阵知识在后续章节中还将用到，且非常关键。

本书各章节组成一个有机的整体，而并非孤立存在。本书路线图中的知识点前后呼应，能为读者提供一些有趣而深刻的洞见。所以，笔者强烈建议读者按顺序阅读全部内容。

代码与网络资源

读者可在 gitee.com/neural_network/neural_network_code 找到本书的样例代码，后续我们将继续维护和扩展这个代码库。笔者在知乎开设了专栏“计算主义”（zhuanlan.zhihu.com/pillgrim），以后将继续在专栏中讨论和分享本书相关主题。鉴于笔者水平有限，错误难免，欢迎读者来信指正，邮箱是 zhangjuefei83@163.com。

致谢

感谢 360 智能工程部“Prophet 机器学习平台”团队的同事：组长陈震、黄斌、林灯博士、孙晓冬、马宗超、吕仁杰、刘韦菠、蔡春蒙、缪路文。孙晓冬为本书第 11 章编写了样例代码，感谢他杰出的工作。

感谢北京图灵文化发展有限公司的陈兴璐女士，她的邀约和鼓励促使本书得以面世，她出色的编辑工作使本书避免了很多问题和缺陷。

感谢中国科学院古脊椎动物与古人类研究所的潘雷博士和北京理工大学生物医学工程系的陈端端教授，她们在参考文献方面给予极大的帮助。

感谢中国科学院古脊椎动物与古人类研究所副研究员朱幼安博士和上海社会科学院哲学研究所钱立卿博士，感谢他们多年来与我在生命演化、混沌系统和计算理论方面的共同兴趣和探讨。

感谢北京自然博物馆标本部副研究员刘迪，蒙他惠允于本书中使用鸟类骨骼与生态类群数据集。

最后，感谢我的家人。



微信连接



回复“深度学习”查看相关书单



微博连接

关注@图灵教育 每日分享IT好书



QQ连接

图灵读者官方群I: 218139230

图灵读者官方群II: 164939616

图灵社区
iTuring.cn

在线出版,电子书,《码农》杂志,图灵访谈

目 录

第一部分 线性模型

第 1 章 逻辑回归	2
1.1 作为一个神经元的逻辑回归	2
1.2 基础向量几何	4
1.2.1 向量	4
1.2.2 向量的和、数乘与零向量	6
1.2.3 向量的内积、模与投影	8
1.2.4 线性空间、基与线性函数	11
1.2.5 直线、超平面与仿射函数	14
1.3 从几何角度理解逻辑回归的能力和局限	17
1.4 实例：根据鸟类骨骼判断生态类群	20
1.5 小结	24
第 2 章 模型评价与损失函数	25
2.1 训练集与测试集	25
2.2 分类模型的评价	26
2.2.1 混淆矩阵	26
2.2.2 正确率	27
2.2.3 查准率	27
2.2.4 查全率	27
2.2.5 ROC 曲线	28
2.3 损失函数	29
2.3.1 K-L 散度与交叉熵	29
2.3.2 最大似然估计	31
2.3.3 从几何角度理解交叉熵损失	33
2.4 小结	35

第 3 章 梯度下降法

3.1 多元函数的微分	36
3.1.1 梯度	37
3.1.2 方向导数	40
3.1.3 偏导数	43
3.1.4 驻点	43
3.1.5 局部极小点	44
3.2 梯度下降法	46
3.2.1 反梯度场	47
3.2.2 梯度下降法	49
3.2.3 梯度下降法的问题	50
3.3 梯度下降法的改进	52
3.3.1 学习率调度	52
3.3.2 冲量法	54
3.3.3 AdaGrad	55
3.3.4 RMSProp	56
3.3.5 Adam	57
3.4 运用梯度下降法训练逻辑回归	59
3.5 梯度下降法训练逻辑回归的 Python 实现	61
3.6 小结	67

第 4 章 超越梯度下降

4.1 矩阵	68
4.1.1 矩阵基础	68
4.1.2 矩阵的逆	71
4.1.3 特征值与特征向量	73
4.1.4 对称矩阵的谱分解	74

4.1.5 奇异值分解	76
4.1.6 二次型	77
4.2 多元函数的局部二阶特性	79
4.2.1 赫森矩阵	79
4.2.2 二阶泰勒展开	79
4.2.3 驻点的类型	82
4.2.4 赫森矩阵的条件数	84
4.3 基于二阶特性的优化	87
4.3.1 牛顿法	87
4.3.2 共轭方向法	92
4.4 运用牛顿法训练逻辑回归	95
4.5 牛顿法训练逻辑回归的 Python 实现	98
4.6 小结	100
第 5 章 正则化	102
5.1 概率论回顾	102
5.1.1 随机变量	102
5.1.2 多元随机变量	105
5.1.3 多元随机变量的期望和协方差矩阵	106
5.1.4 样本均值和样本协方差矩阵	106
5.1.5 主成分	108
5.1.6 正态分布	111
5.2 模型自由度与偏置-方差权衡	115
5.2.1 最小二乘线性回归	116
5.2.2 模型自由度	118
5.2.3 偏置-方差权衡	119
5.3 正则化	122
5.3.1 岭回归与 L_2 正则化	122
5.3.2 L_2 正则化的贝叶斯视角	125
5.3.3 L_1 正则化	126
5.4 过拟合与欠拟合	127
5.5 运用 L_2 正则化训练逻辑回归	130
5.6 运用 L_2 正则化训练逻辑回归的 Python 实现	132
5.7 小结	135

第二部分 神经网络

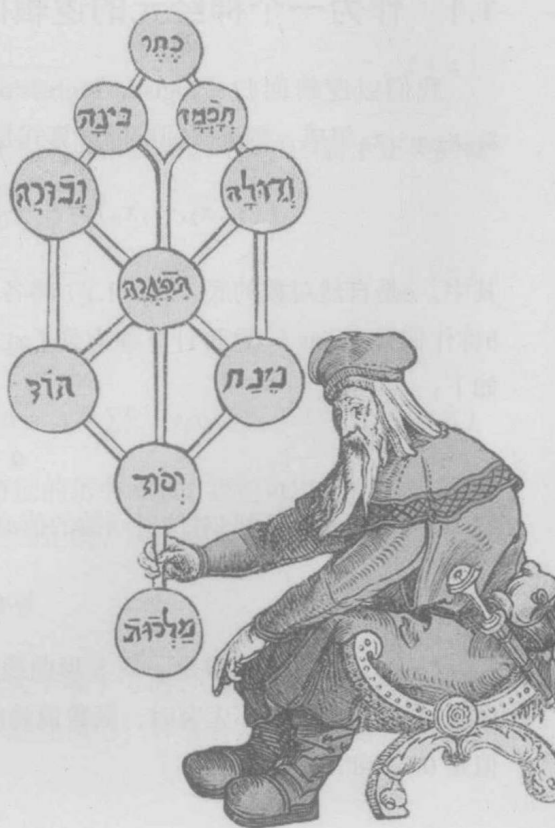
第 6 章 神经网络	138
6.1 合作的神经元	138
6.2 多层全连接神经网络	142
6.3 激活函数	145
6.3.1 Linear	145
6.3.2 Logistic	146
6.3.3 Tanh	148
6.3.4 ReLU	150
6.3.5 Leaky ReLU 以及 PReLU	151
6.3.6 SoftPlus	153
6.4 多分类与 SoftMax	154
6.5 小结	157
第 7 章 反向传播	158
7.1 映射	158
7.1.1 仿射映射	158
7.1.2 雅可比矩阵	159
7.1.3 链式法则	160
7.2 反向传播	162
7.2.1 网络的符号表示	162
7.2.2 原理	163
7.2.3 实现	166
7.3 相关问题	169
7.3.1 计算量	169
7.3.2 梯度消失	170
7.3.3 正则化	170
7.3.4 权值初始化	170
7.3.5 提前停止	171
7.4 多层全连接神经网络的 Python 实现	173
7.5 小结	181
第 8 章 计算图	183
8.1 计算图模型	183
8.1.1 简介	183
8.1.2 多层全连接神经网络的计算图	187

8.1.3 其他神经网络结构的计算图	188	9.3.5 数据增强	239
8.2 自动求导	190	9.4 小结	239
8.3 自动求导的实现	192	第 10 章 经典 CNN	241
8.4 计算图的 Python 实现	195	10.1 LeNet-5	241
8.5 小结	214	10.2 AlexNet	245
第 9 章 卷积神经网络	215	10.3 VGGNet	248
9.1 卷积	215	10.4 GoogLeNet	251
9.1.1 一元函数的卷积	215	10.5 ResNet	255
9.1.2 多元函数的卷积	219	10.6 小结	257
9.1.3 滤波器	223	第 11 章 TensorFlow 实例	258
9.2 卷积神经网络的组件	228	11.1 多分类逻辑回归	258
9.2.1 卷积层	228	11.2 多层全连接神经网络	266
9.2.2 激活层	230	11.3 LeNet-5	269
9.2.3 池化层	231	11.4 AlexNet	273
9.2.4 全连接层	233	11.5 VGG16	277
9.2.5 跳跃连接	234	11.6 小结	280
9.3 深度学习的正则化方法	236	附录 A CNN 与元胞自动机	281
9.3.1 权值衰减	236	参考文献	311
9.3.2 Dropout	237		
9.3.3 权值初始化	237		
9.3.4 批标准化	238		

第一部分

线性模型

- 第1章 逻辑回归
- 第2章 模型评价与损失函数
- 第3章 梯度下降法
- 第4章 超越梯度下降
- 第5章 正则化



统计学习领域的泰斗 Trevor Hastie 曾经说过：“每一个机器学习研究者都应该像熟悉自家后院那样熟悉线性模型。”线性模型是许多非线性模型的基础，它良好的数学结构能够为理解非线性模型提供深刻的洞见。人工神经网络的基本组成单元——神经元，正是线性模型。本章将介绍一种常用的线性模型——逻辑回归：首先引入逻辑回归的定义，之后回顾必要的线性代数知识，并从几何角度阐述逻辑回归及所有线性模型的能力局限。

1.1 作为一个神经元的逻辑回归

我们以逻辑回归（logistic regression）为例讨论线性模型。假如样本由 n 个数值型特征 $x_1, x_2, \dots, x_n$ 组成，则逻辑回归的计算式是：

$$f(x_1, x_2, \dots, x_n) = \frac{1}{1 + e^{-(b + w_1 x_1 + w_2 x_2 + \dots + w_n x_n)}} = \frac{1}{1 + e^{-(b + \sum_{i=1}^n w_i x_i)}} \quad (1.1)$$

其中， e 是自然对数的底。式（1.1）将各个特征 x_i 乘以对应的权重系数 w_i 后相加，之后再加上 b ， b 称作偏置（bias）。这种计算称为关于 $x_1, x_2, \dots, x_n$ 的仿射函数（affine function），仿射函数的值 a 如下：

$$a = b + \sum_{i=1}^n w_i x_i \quad (1.2)$$

接下来，逻辑回归对仿射函数的值 a 施加 Logistic 函数：

$$\text{Logistic}(a) = \frac{1}{1 + e^{-a}} \quad (1.3)$$

Logistic 函数的图像是一条 S 形曲线（也称 sigmoid 曲线）。当 a 趋向于负无穷时，函数值趋向于 0；当 a 趋向于正无穷时，函数值趋向于 1。Logistic 的函数值在 $(0, 1)$ 内，当 $a = 0$ 时，函数值是 0.5，如图 1-1 所示。