

把PaddlePaddle框架的函数与深度学习、
神经网络等知识结合起来讲解

通过丰富的案例，由浅入深地介绍了PaddlePaddle在典型场景中的应用实践



PaddlePaddle 与深度学习应用实战

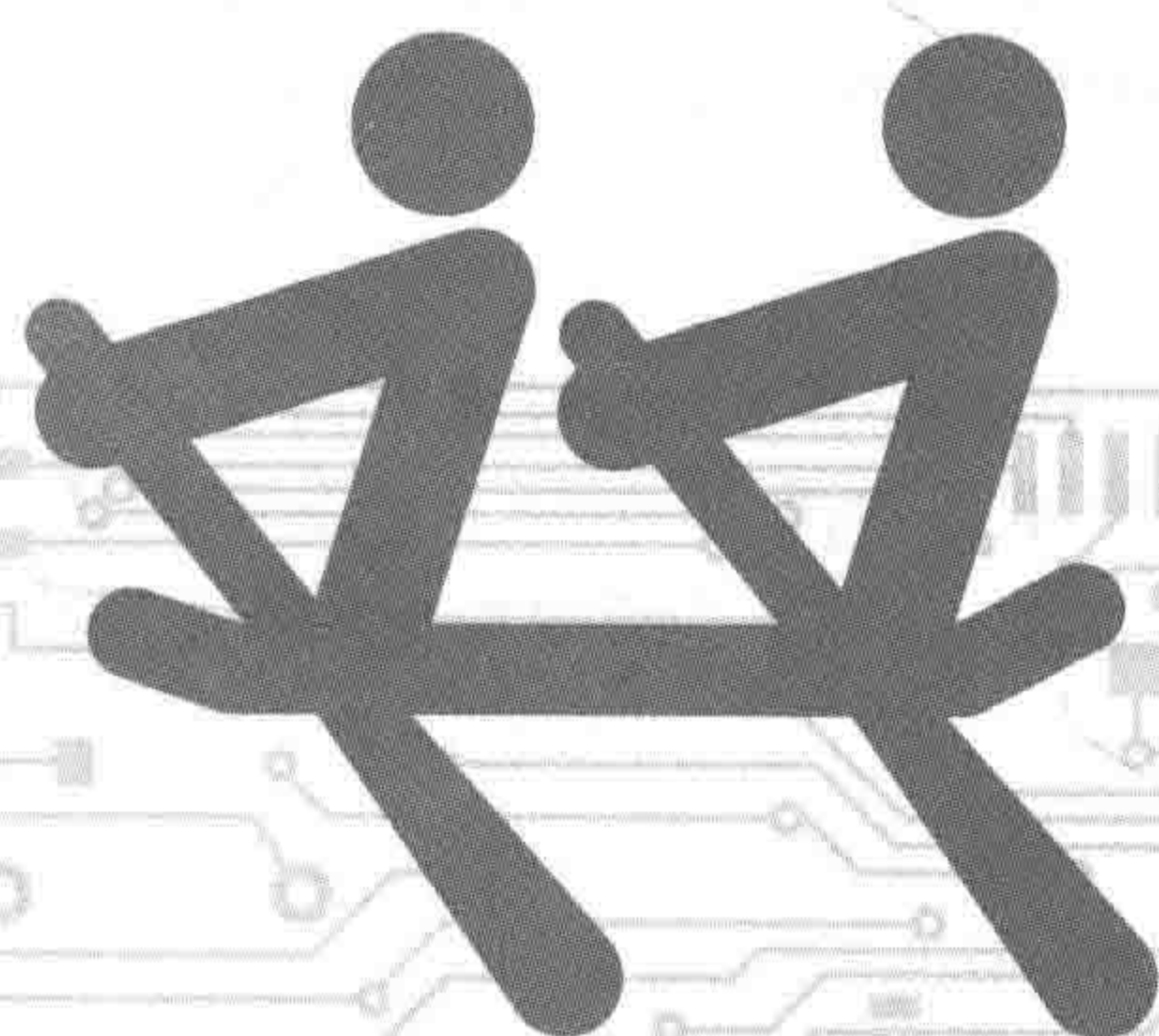
程天恒 编著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



PaddlePaddle

与深度学习应用实战

程天恒 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

深度学习是目前人工智能研究中前沿、有效的一项技术，主要通过构建深度神经网络解决视觉、自然语言处理、语音识别等诸多领域的问题。百度在 2016 年发布了国内首个开源深度学习框架 PaddlePaddle，简化了深度学习算法的实现步骤，提供了灵活、易用的接口，同时支持分布式训练。

本书由简单的例子引入深度学习和 PaddlePaddle 框架，介绍了 PaddlePaddle 的安装、测试与基本使用，并结合 PaddlePaddle 接口介绍深度学习的基础知识，包括常用的神经网络和算法。最后，通过一系列深度学习项目实例介绍 PaddlePaddle 在各种场景和问题中的应用，让读者由浅至深地理解并运用深度学习解决实际问题。

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

PaddlePaddle 与深度学习应用实战 / 程天恒编著. —北京：电子工业出版社，2018.7
ISBN 978-7-121-34247-9

I. ①P… II. ①程… III. ①学习系统 IV. ①TP273

中国版本图书馆 CIP 数据核字(2018)第 106118 号

责任编辑：安 娜

印 刷：三河市双峰印刷装订有限公司

装 订：三河市双峰印刷装订有限公司

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：787×980 1/16 印张：14.5 字数：276 千字

版 次：2018 年 7 月第 1 版

印 次：2018 年 7 月第 1 次印刷

定 价：65.00 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 zlts@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。

前言

深度学习是当下十分火热的技术之一，在大数据和大规模高速并行计算的帮助下，深度神经网络在各大领域开始发挥出巨大威力。“刷脸”解锁、自动驾驶、机器翻译、图像识别，这些技术已经扎根于我们的生活之中。有人说“21 世纪是人工智能的世纪”，我很赞同这个观点，尤其近几年，出现了像 AlphaGo、Apollo 自动驾驶这样的技术浪潮。在学术界，机器学习相关会议和期刊投稿数逐年翻倍，越来越多的学者转向人工智能领域，或是研究机器学习最基本的理论；或是不断追求更好的方法和模型来解决计算机视觉、自然语言处理及数据挖掘；或是开始寻找新方向，走出一条“AI+”之路，将深度学习应用到更广阔的领域，如医疗、零售等。很多人认为这些都是科学理论，离我们还很遥远，其实不然，人工智能应用已经在不断靠近并改善我们的生活。

在这样一个浪潮趋势下，可能越来越多的人会投身于人工智能领域，但很多资料和文献门槛过高，丰富的数学理论知识，让很多人望而却步。因此，借着这个机会，我想通过这本书把我所学分享给大家，让我们共同学习，共同创造，共同为人工智能的发展贡献一份力量。

本书内容

本书共 9 章，首先从基础知识入手，将 PaddlePaddle 框架的函数与深度学习知识相结合，带领读者灵活搭建神经网络，并选择合适的优化算法和激活函数，训练神经网络。

结合目前前沿的研究及样例代码，帮助读者加深对卷积神经网络和循环神经网络的理解。

接着用丰富的案例，如图像识别、图像描述及聊天机器人等，通过实例讲解如何将 PaddlePaddle 框架应用到实际应用中。

最后，介绍了对抗网络（GAN）及强化学习的基本思想和应用，通过解读对抗网络（GAN）的官方代码，帮助读者理解 GAN。

本书面向的读者

对深度学习感兴趣的初学者。对于初学者，本书将 PaddlePaddle 框架和深度学习的基本概念和基本原理相结合，在学习理论知识的同时掌握了一个高效的深度学习框架。

人工智能领域的研究者及从业者。对于从业者，本书更是一本工具书，读者可以通过阅读本书学习 PaddlePaddle 框架，利用丰富的实例和代码快速上手，并将 PaddlePaddle 框架运用到自己的工作和研究中。

致谢

感谢百度公司 PaddlePaddle 团队，开发出一款高效、易用、易学的深度学习框架 PaddlePaddle，并成为国内首个开源深度学习平台。

感谢百度公司 PaddlePaddle 开发者和开源社区的朋友，能够快速回复我的每一个 GitHub Issue，并耐心指导我解决一些问题。

感谢本书所引用著作和论文的作者们，你们的学习成果为我打开了通往人工智能新世界的大门，通过学习你们在这些领域的知识，我对深度学习的理论知识及其应用都得到了提升和加强。

感谢本书的每一位读者，你们的存在是对我最大的支持和鼓励。当然，如果在阅读过程中发现了一些错误或者疑问，十分欢迎与您交流沟通，我的邮箱是 paddle_readers@126.com。由于本人能力有限，因此书中可能存在一些不恰当的表述或者遗漏，还请多多包涵。

程天恒

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。

- ◎ **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，与其他读者一同学习交流。

页面入口：<http://www.broadview.com.cn/34247>



目录

第 1 章	深度学习简介.....	1
1.1	初见	1
1.2	机器学习	1
1.3	神经网络	3
1.4	深度学习介绍	7
1.5	深度学习应用	8
1.6	深度学习框架	12
1.7	深度学习的未来	15
第 2 章	PaddlePaddle 简介.....	16
2.1	安装 PaddlePaddle.....	16
2.2	测试 PaddlePaddle.....	29
第 3 章	初探手写数字识别	31
第 4 章	PaddlePaddle 基本用法	44
4.1	数据准备	44
4.2	原始数据读取及预处理	44
4.3	PaddlePaddle 训练数据.....	46
4.4	模型配置	52
4.5	激活函数	58
4.6	优化方法	64
4.7	损失函数	72
4.8	均方损失函数	73

4.9	交叉熵损失函数	73
4.10	Huber 损失函数	74
4.11	CRF 损失函数	74
4.12	CTC 损失函数	75
4.13	反向传播算法	75
第 5 章	卷积神经网络	78
5.1	卷积神经网络	78
5.2	实例学习	87
5.3	拓展	112
第 6 章	循环神经网络	118
6.1	RNN 简介	118
6.2	双向循环神经网络	121
6.3	循环神经网络使用场景	127
6.4	预测 sin 函数序列	129
6.5	拓展	134
第 7 章	PaddlePaddle 实战	136
7.1	自编码器	136
7.2	PaddlePaddle 实现自编码器	137
7.3	实战 OCR 识别（一）	140
7.4	实战 OCR 识别（二）	150
7.5	情感分析	164
7.6	Seq2Seq 及其应用	172
7.7	实现	178
7.8	Image Caption	194
第 8 章	深度学习新星：生成对抗网络 GAN	208
8.1	生成对抗网络（GAN）	208
8.2	GAN 的其他应用	213
第 9 章	强化学习与 AlphaGo	216

深度学习简介

1.1 初见

2016年4月，AlphaGo 与韩国顶尖棋手李世乭对战，最终 AlphaGo 以 4 胜 1 平战胜了人类，人工智能也因此成为科技界乃至整个社会的热点话题。2017 年，CMU 大学开发的“冷扑大师”——Libratus 无限德州扑克人工智能系统又一次颠覆了 AI 在人们心中的地位。究竟是什么支撑了人工智能，使其有如此高的智商与巨大的威力？

“人工智能”可分为“人工”和“智能”两部分，由人工构建智能化系统，使其具有人类一样的智能性，如思考与学习能力等。这一概念早在计算机还未普及的时代就已由大师图灵提出，后来诞生了一系列科幻小说等。现在，这一研究发展迅猛，在很多领域已开始普及应用。

说起当下的人工智能，就不得不提它背后的算法支持，即深度学习，以及深度学习的基础——机器学习。

1.2 机器学习

不知不觉，机器学习早已进入我们生活的各个角落。举个生活中的例子，我们每个月总会收到几封广告之类的垃圾邮件，有些人甚至每天都会收到垃圾邮件，但我们使用的邮箱服务网站一般都会帮我们将垃圾邮件筛选出来。因此，你可能会问，是不是有客服之类的人在后台帮我们监视？其实并不是，没有这么多的人力资源来完成这件事。垃圾邮件处理是计算机程序帮我们完成的。它们的工作就是收到一封邮件后，查看邮件内容，然后判断是否为垃圾邮件。这和机器学习又有什么关系呢？这就关系到我们如何获得这个分类程序了。要想得到这个垃圾邮件分类程序，就需要一些邮件数据（既含有垃圾邮件，又含有普通邮件）以及邮件的标记（标记对应邮件是否为垃圾邮件），有了这些数据后，我们就可以通过算法来构建一个模型。例如，我们可以用获得的数据画一条线性回归曲线，有了新邮件之后，便可以通过这条曲线判断是否

为垃圾邮件，我们的模型通过“吃掉”数据来学习判断垃圾邮件，这就是机器学习。

综合来看，机器学习就是通过数据提取特征，然后结合学习算法来构建一个模型，这个模型就是用来分类垃圾邮件的引擎，如图 1.1 所示。

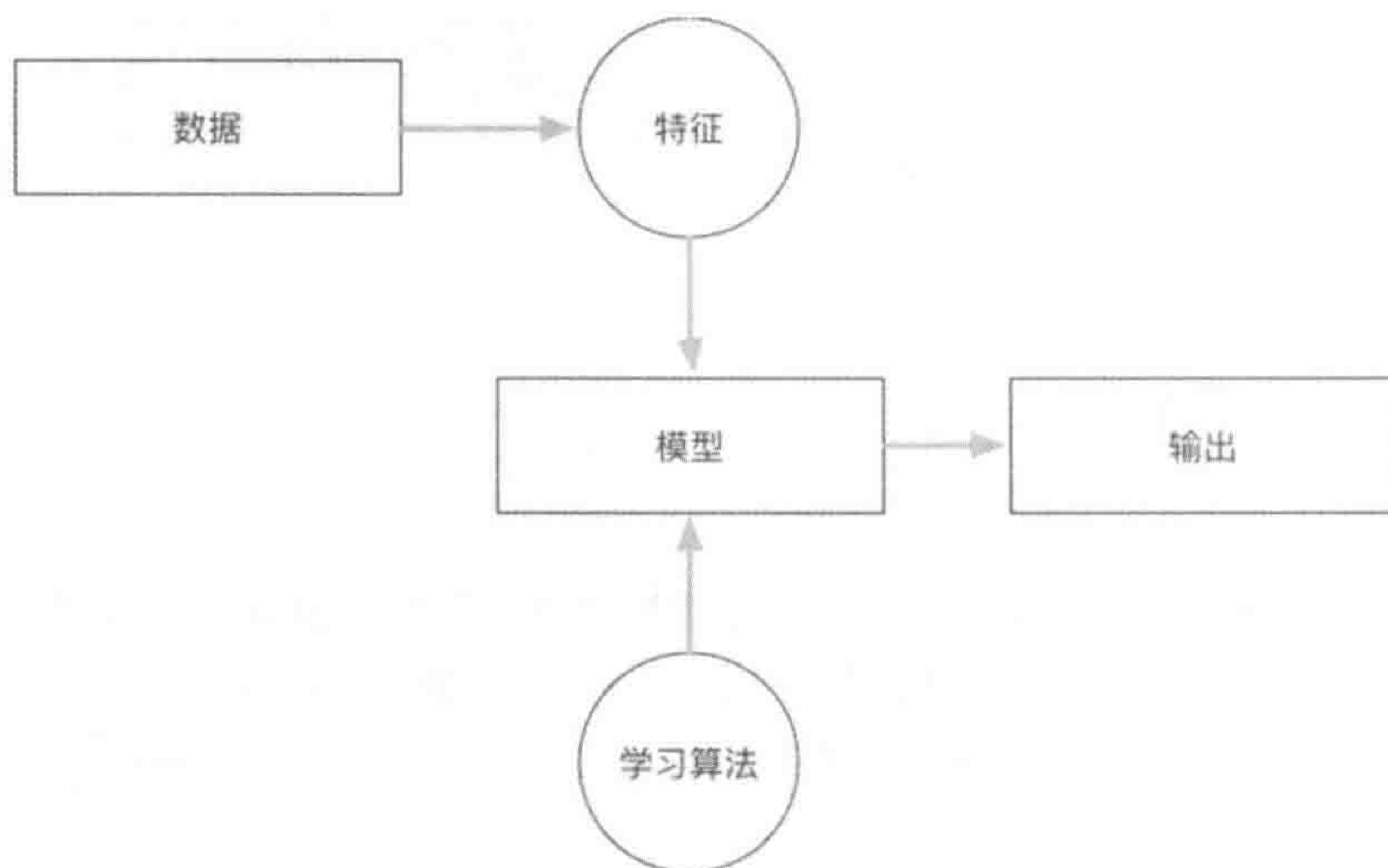


图 1.1

机器学习是一个庞大的体系，其中囊括了众多学习算法，如贝叶斯分类、决策树、支持向量机等，此外，机器学习与数学联系紧密，尤其是概率论和统计学。

分类与回归。在生活中，很多问题可以看作分类（Classification）问题或者回归（Regression）问题，我们的大脑也无时无刻不在处理这些问题。例如，看书时，我们需要知道这是一本什么书，是小说还是工具书。去动物园游玩时，我们每看到一种动物都会利用所看见的内容，对眼前的动物进行分类。一般对于离散的问题，可以采取分类的方法，而连续的问题，更多的是使用回归，如股票预测、温度预测等。

监督学习和无监督学习。机器学习在训练中可以分为监督学习（Supervised Learning）和无监督学习（Unsupervised Learning）两种。监督学习是指训练模型时，我们除了提供输入数据（ X ），还需要提供一个标准结果（Label），模型通过读取输入数据 X 进行预测得到一个预测结果 Y ，然后对比 Label 和 Y 的差异，优化模型参数。监督学习可以想象为我们给机器提供一张图像，然后告诉它图像里有什么物体，机器就利用这种方式学习识别图像，随着样本数量增多，训练迭代次数增长到一定数量之后，机器便能自主地识别物体，而不再需要标签。在机器学习中，支持向量机、贝叶斯等算法都是基于监督学习的。而无监督学习与监督学习的最大区别是，无监督学习不存在标签（Label），全靠输入去寻找特征，然后归类。聚类算法就是一种无监督学习，即利用数据本身的特征进行归类，将相似的数据归为一类，将差别较大的数据分

开。当然，机器学习中的无监督学习不只有聚类算法，还有很多其他的算法，在此就不一一介绍了。

1.3 神经网络

神经网络（Neural Network）是机器学习的一个分支，起源于人们对生物体神经网络的认知。生物神经网络由神经元、突触等结构组成，大量的神经元通过无数的突触连接可以构成一个大规模神经网络，能够处理人的思维和记忆，如图 1.2 所示。

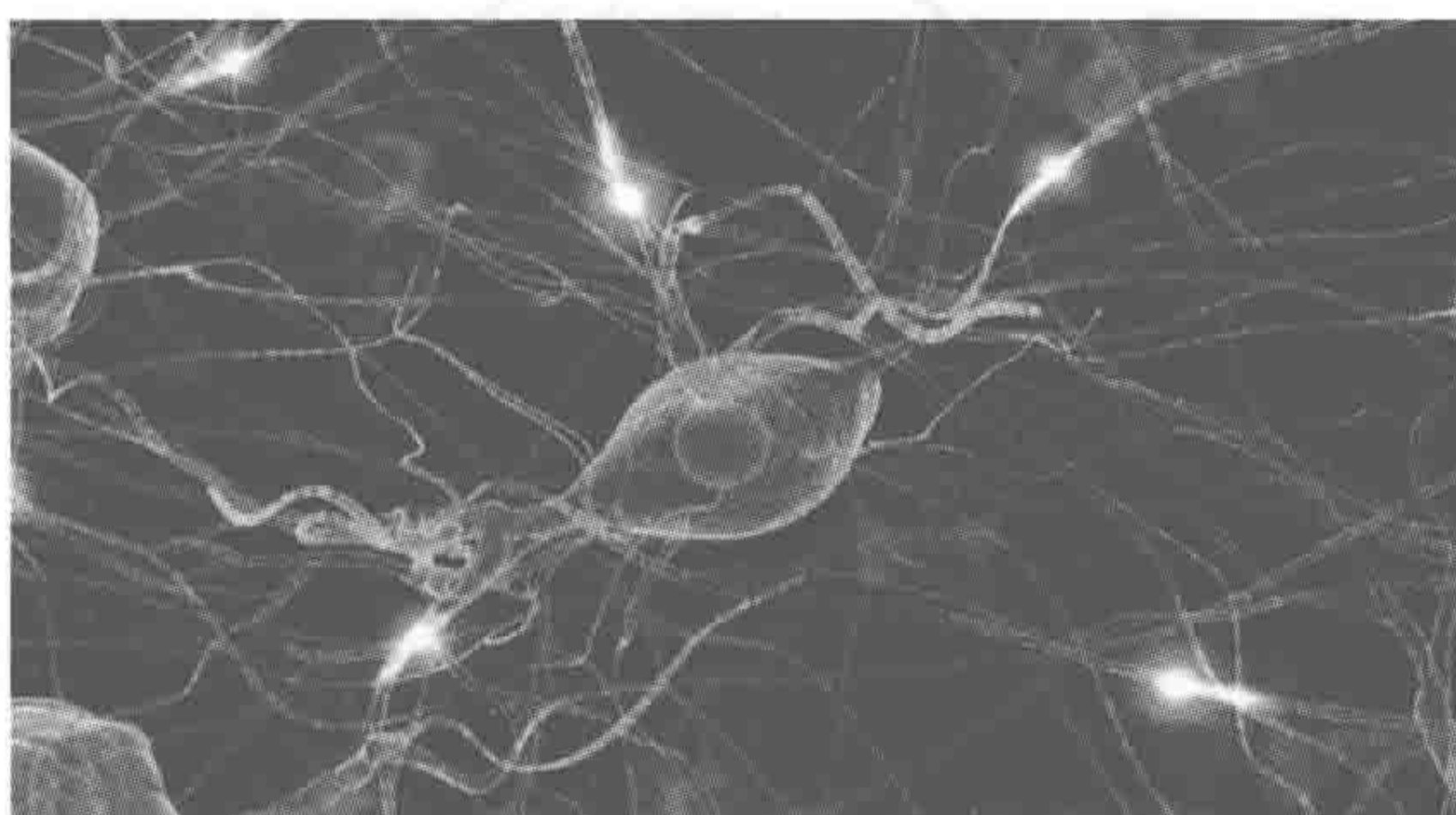


图 1.2

人们通过模仿生物神经网络的工作原理构建了人工神经网络（Artificial Neural Network）。与生物神经网络相同，人工神经网络中也是先建立一些神经元模型，早期人们称之为感知机（Perception），然后将所有的神经元模型连接起来，形成网络。

人脑神经元结构如图 1.3 所示。

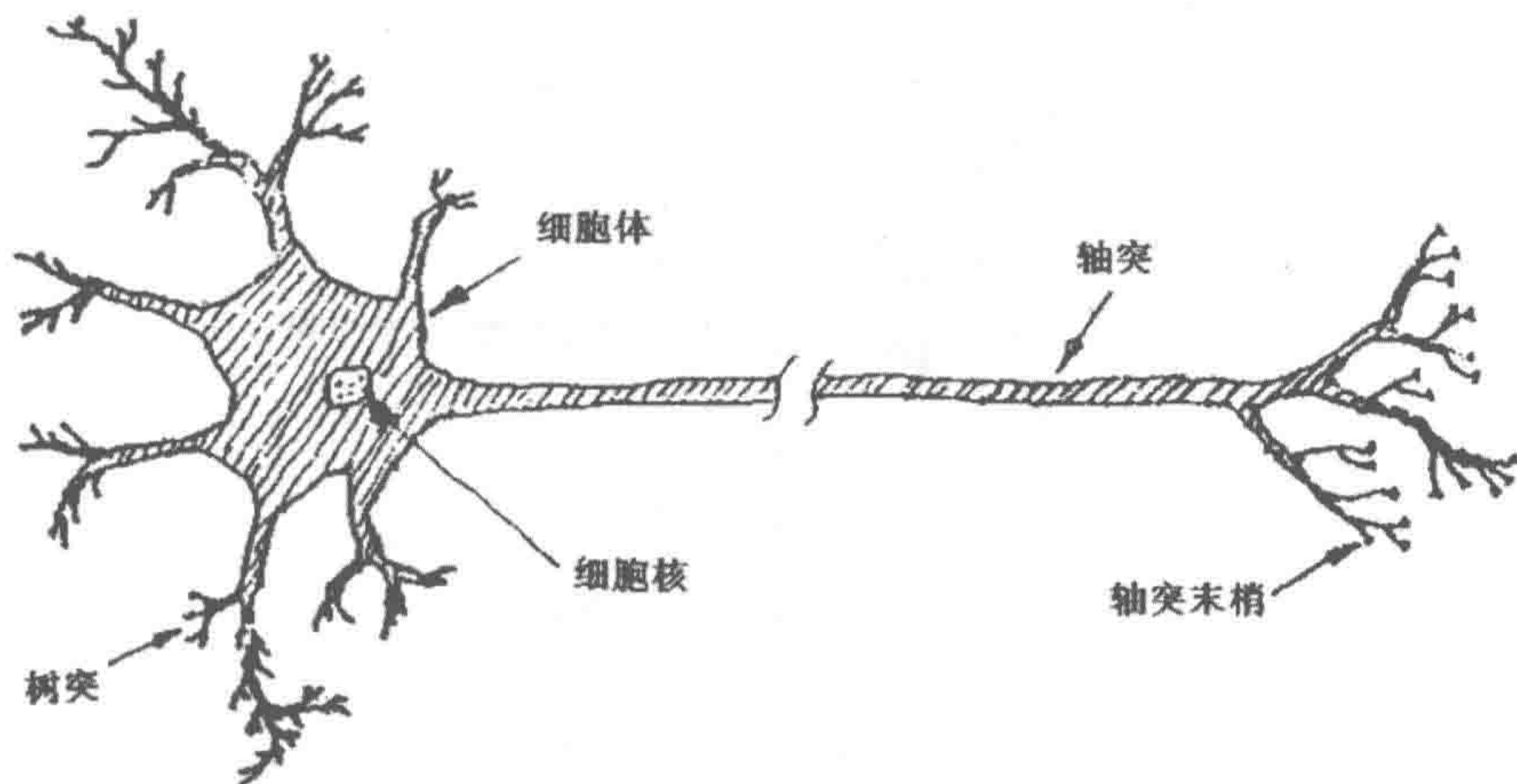


图 1.3

人工神经元结构如图 1.4 所示。

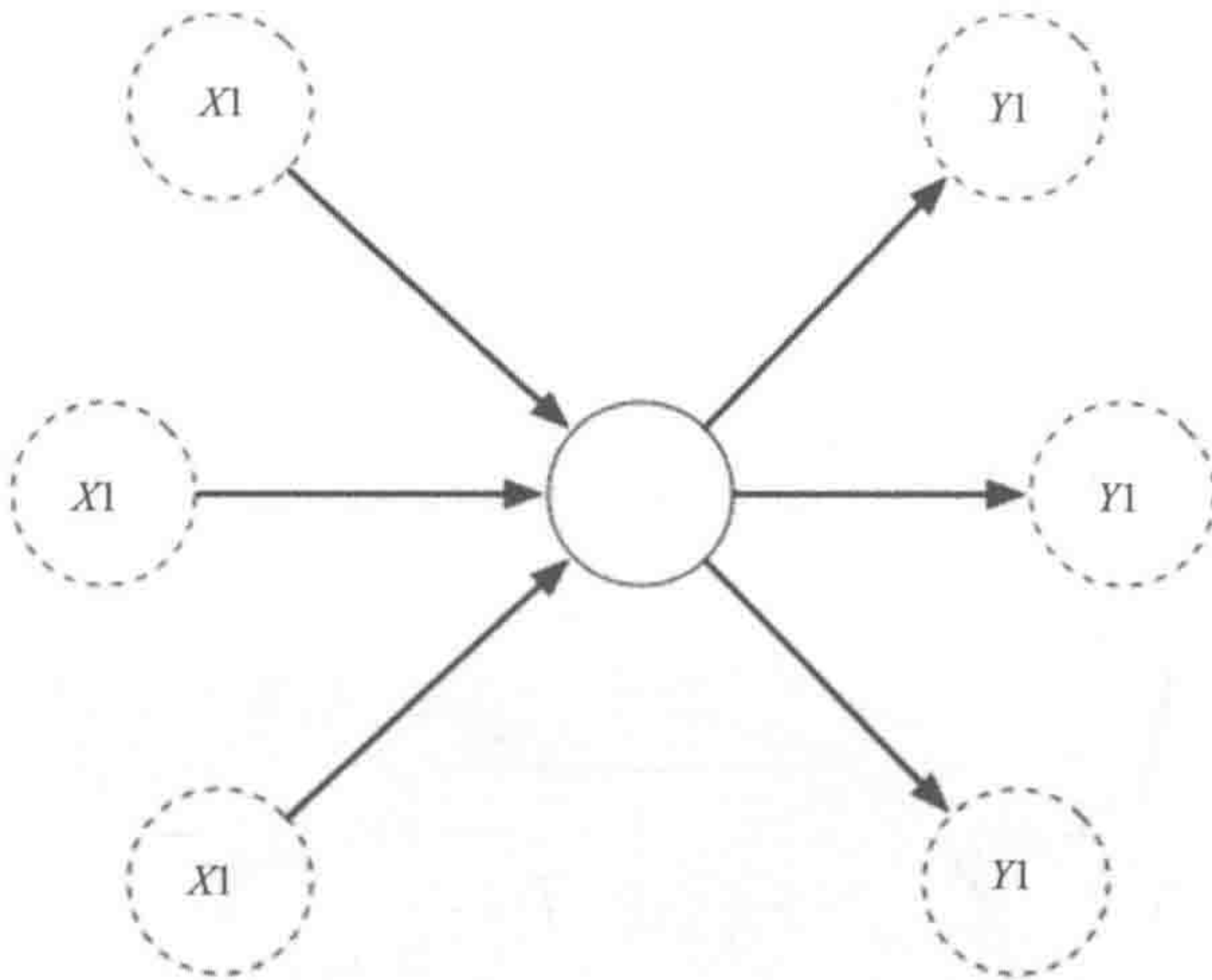


图 1.4

人工神经元获取其他神经元提供的输入，对其加权求和，然后利用特定的激活函数去处理求和结果以得到输出，然后将输出传递给下一层神经元。

一个神经元的作用只是一个简单线性函数，而我们的神经网络就是将多个神经元组合起来形成一层网络。单层神经网络支持高维度的输入和输出，同时也可以添加非线性函数来激活。

在深入学习神经网络之前，让我们先来了解单个神经元的工作原理。

设定输入为 $[x_1, x_2, x_3, x_4]$ ，其对应的权重为 $[w_1, w_2, w_3, w_4]$ ，如图 1.5 所示。

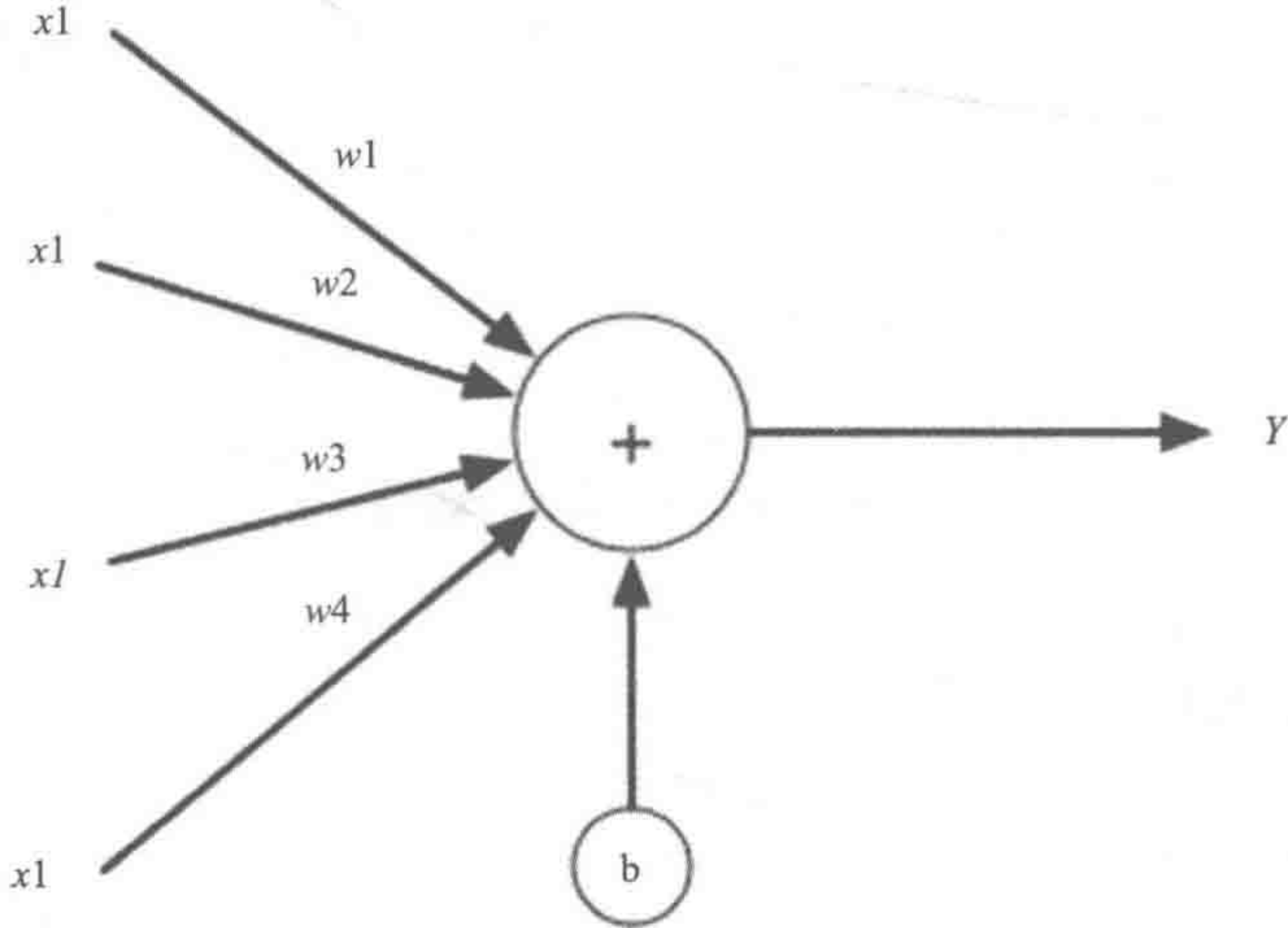


图 1.5

加权和：

$$S = x_1\omega_1 + x_2\omega_2 + x_3\omega_3 + x_4\omega_4$$

然后经过线性或者非线性函数进行激活：

$$Y = f(S + b)$$

b 为偏置变量。

当 $[w_1, w_2, w_3, w_4]$ 为向量时，我们得到的输出是一个向量而不是单个值。

我们可以将多个神经元组成一层神经元层（Layer），如图 1.6 所示。

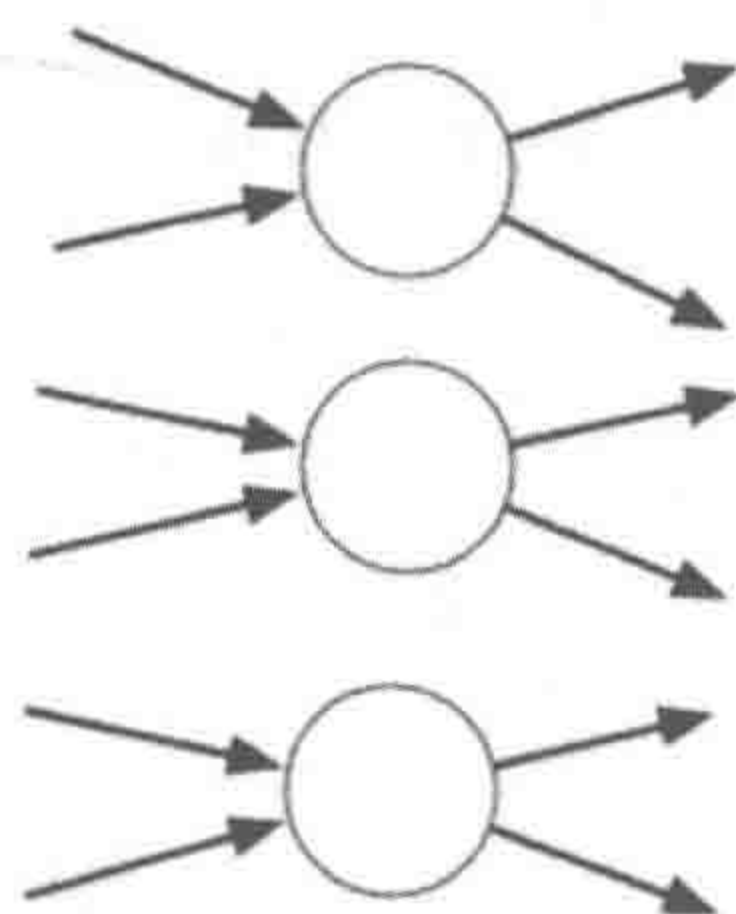


图 1.6

如果这样的神经元层有多层，并且每层之间相互连接，那么便可以构造出一个神经网络，如图 1.7 所示。

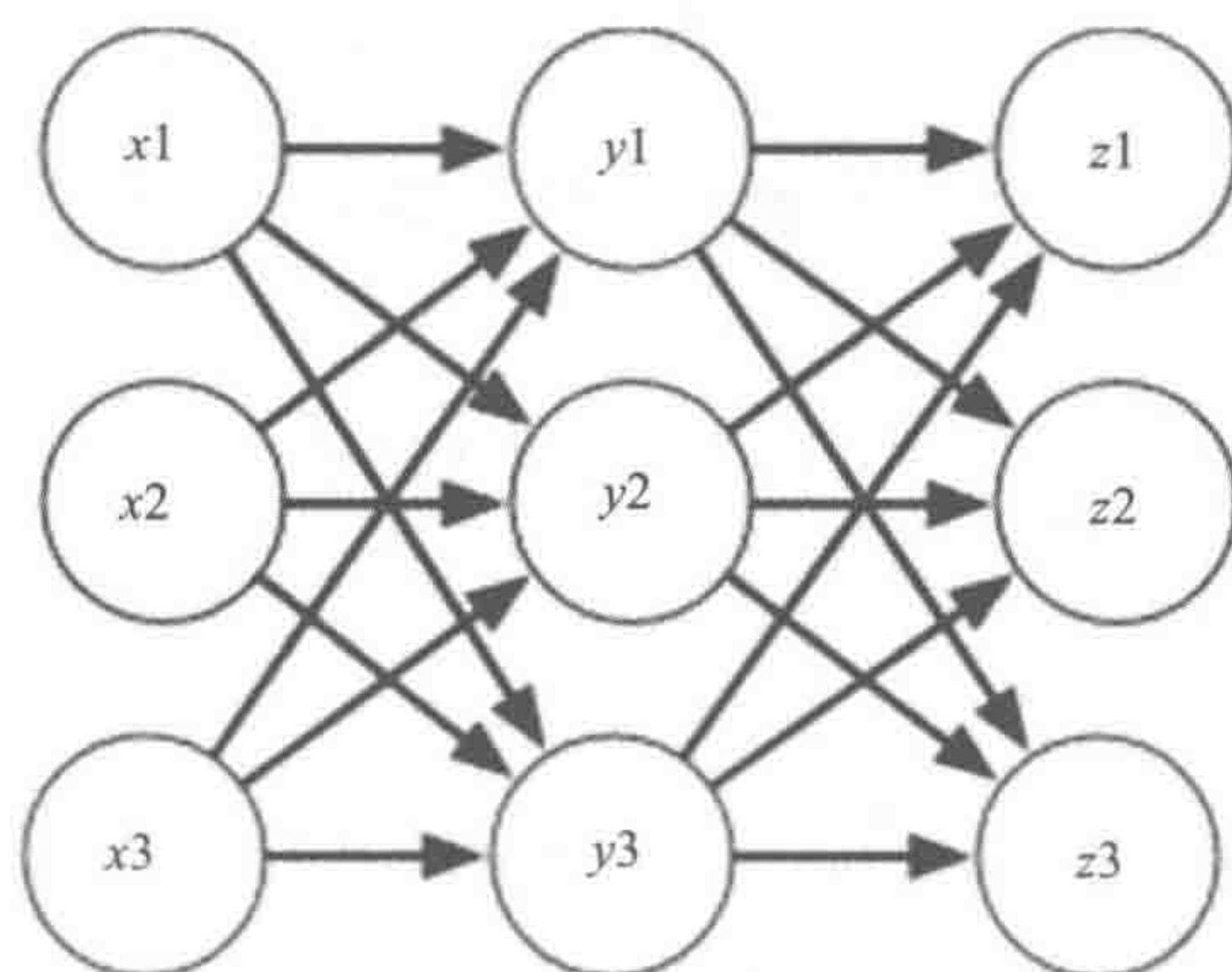


图 1.7

图 1.7 就是简单的多层神经网络（也称为多层感知机），即一个网络中包含多层感知机（神经元）。

单层感知机模型其实早在几十年前就已经问世了，当时人们利用这种模型来解决简单的线性分类问题，但当时人们对感知机的认识还仅局限在单层感知机，多层感知机计算过于复杂。使用感知机解决异或问题（XOR）时，分类出现了问题，单层感知机模型根本无法解决这个问题，于是神经网络的发展就在那个时代停滞不前，神经网络与人工智能的研究开始进入了寒冬期。

直到 1986 年，Geoffery Hinton 等人提出了反向传播算法（Backpropagation），解决了多层感知机的优化问题，同时也使得多层感知机能够自我训练。有了多层感知机，异或等线性不可分问题便迎刃而解。

直到今天，神经网络依然依赖于反向传播进行训练，只是网络结构和模型越来越复杂。

简单的神经网络有一层输入层、一层隐藏层（Hidden Layer）和一层输出层，如图 1.8 所示。

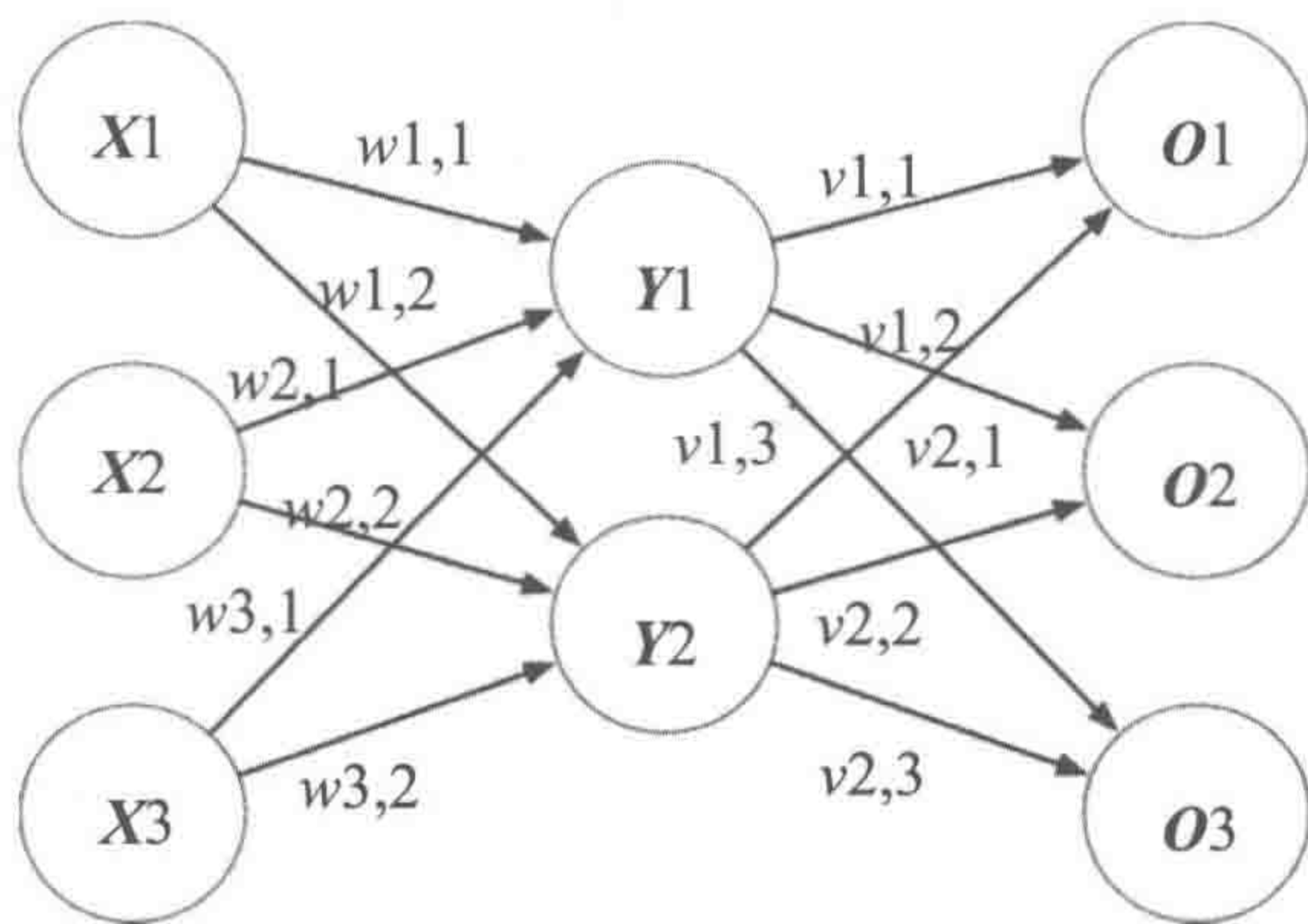


图 1.8

其中 X 为输入值向量， Y 为中间隐层向量， O 为输出层向量，我们可以按照单个神经元的计算方式来计算每一步的输出：

$$Y_1 = f(x_1\omega_{11} + x_2\omega_{21} + x_3\omega_{31} + b_1)$$

$$Y_2 = f(x_1\omega_{12} + x_2\omega_{22} + x_3\omega_{32} + b_2)$$

输出层 O 的结果：

$$O_1 = \sigma(Y_1v_{11} + Y_2v_{21} + b_1)$$

$$O_2 = \sigma(Y_1v_{12} + Y_2v_{22} + b_2)$$

$$O_3 = \sigma(Y_1v_{13} + Y_2v_{23} + b_3)$$

多层神经网络足以拟合任何函数，我们可以通过增加深度和宽度来提高模型的健壮性和预测的准确率。

深度学习和神经网络又有什么关系呢？

1.4 深度学习介绍

前面介绍了机器学习和简单的神经网络，以及机器学习的原理。深度学习是在机器学习之上发展而来的。最初，我们使用简单的感知机来完成非线性分类问题，之后，出现了非线性激活函数，感知机层数也逐渐增加，人工神经网络（Artificial Neural Network）便被创造出来了。深度学习就是利用深层神经网络去学习，完成类似于分类这样的任务。早期，我们通常使用多层神经网络来构建模型，后来又诞生了卷积神经网络（Convolutional Neural Networks）、循环神经网络（Recurrent Neural Networks）等模型，让整个深度学习体系变得更加完整。

传统的机器学习算法对原始数据处理能力不强，无法获得内部深层次的特征，在数十年的研究中，人们一直在寻找一些方法对数据进行处理，将其内部信息以一种方式表现出来，例如，一些矩阵或者向量，通过一些变换，用复杂的非线性函数实现图像的分类和识别等。

深度学习是建立在人工神经网络之上的，对于一般的应用，我们会采用监督学习的方式来训练我们的模型。例如，我们常常说的图像分类任务，就是将相机拍出的照片转化为图像矩阵，通过我们预先构建的神经网络模型来进行分类。在没有预先训练的情况下，模型很难预测输入的图像属于哪个类别，在监督学习中，我们会在输入数据的同时告诉它这是哪一个类别，当它得到自己的预测之后，与我们提供的标准类别进行对比，然后去调整模型的一些参数与设置，这个对比就实现了一种监督的机制。在监督学习中，我们会去关注模型预测的损失，也就是预测的结果和标准结果的差距。

深度学习的基础是神经网络，最简单的模型就是含有输入输出层以及一个中间的隐藏层（Hidden Layer），这些都可以用简单的感知机来实现。对于复杂的网络，可能还会有很多卷积神经网络（Convolutional Neural Networks）或者循环神经网络（Recurrent Neural Networks），此外还有各种各样的模型。尽管每一种网络结构都不一样，功能也大不相同，但其计算原理和训练方法都是相通的。

现在的深度学习依赖于三个步骤：前向传播（Forward）、损失计算（Loss）和反向传播（Backward）。

前向传播是输入的数据逐层通过模型，计算每一层的输出作为下一层的输入，直到最后一层输出结果。

损失计算利用预先提供的标签和网络的输出进行对比，结合我们定义的损失计算函数计算误差 loss。

完成损失计算后，就可以利用得到的误差和梯度下降算法（SGD）从网络的输出反向传递误差，并优化每一层的参数，直到输入层。这样一个过程就完成了一次模型训练。

从输入层输入数据，不断向前进行计算，直到输出层输出结果，然后计算损失，接着沿着反方向传递误差和梯度，进行参数修正。

深度学习核心算法就是反向传播算法和梯度下降算法，利用梯度下降算法可以优化我们的模型参数，梯度下降利用迭代的方式进行最优求解。而反向传播算法用于传递误差，将误差逐层向输入层传递，这样，每一层都可以使用梯度下降算法进行优化，最终，一次迭代，整个网络都会得到更新。

深度学习的大门正逐渐向我们敞开，研究论文层出不穷，尤其是图像、语音领域，基于深度学习的应用已经大规模出现在我们生活中了。在这个人工智能的时代，我们应当成为一名先驱者，引领人工智能的发展。

1.5 深度学习应用

1.5.1 图像识别与分类

深度学习的发展并不是一气呵成的，其间经历了很多转折和突破，其中，转折较大的一次就是 ImageNet 的出现。ImageNet 是由斯坦福大学人工智能科学家李飞飞等发起的一个大规模图像数据集项目，拥有目前世界上最大的图像识别数据库，如图 1.9 所示。



图 1.9

为什么说 ImageNet 的出现给深度学习领域带来了一次转折呢？对于深度学习任务来说，大量的数据集是一种必备条件，而当时的环境很难有足够的图像数据用来训练，ImageNet 项目提

供了大量带有标注图像的数据，从2010年至2017年，每年都会举办图像识别评测比赛，在这一驱动下，很多实验室开始了图像分类的研究，不断有新的方法出现来刷新识别记录。卷积神经网络在这个阶段飞速发展，每年都会有新的各种各样的方法，网络模型越来越深，结构也越来越复杂。同时，ImageNet除分类识别任务外，还包括检测和分割等评测比赛。

图像分类主要在于分析输入的图像，对图像中出现的物体进行分类或者识别。例如，大家所熟知的MNIST手写数字，识别任务就是给出图像中到底是什么数字。ImageNet提供了大量的分类数据集，到目前为止，已有14000000多张图像，共标注了20000多个类别。

1.5.2 图像检测

图像检测（Object Detection）和图像分类识别任务不同，图像分类识别任务更倾向于知道图片里的东西是什么，而检测倾向于知道东西在哪里，确定物体的位置。

检测任务主要用来确定物体的位置，而位置信息一般利用一个 Bounding Box 来描述。Bounding Box 通常由四部分组成，即 x 坐标、 y 坐标、宽度以及高度，有时还会加上旋转角度，如图1.10所示。

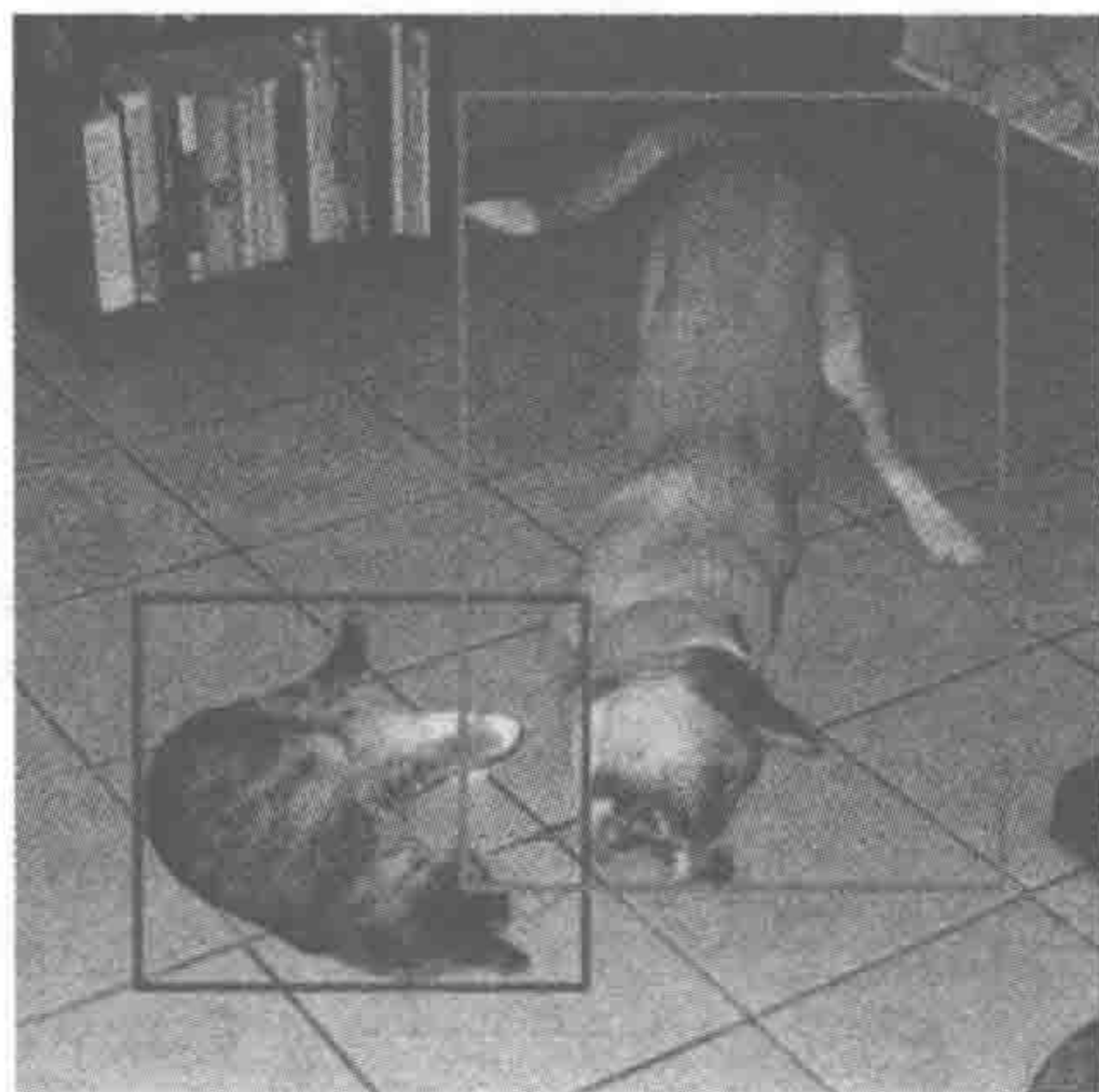


图 1.10

图像检测是计算机视觉的一个基础方向，在复杂场景下，识别任务也需要依赖于检测任务，先检测物体的位置和范围，然后在特定区域进行识别任务。现在，图像检测出现了 SSD、YOLO 和 Faster-RCNN 等出色的深度学习方法，同时，ImageNet 每年都会举办 Object Detection 比赛（已归入 Kaggle）。

1.5.3 图像分割

图像分割（Image Segmentation）可以将图像分割成多个版块。在传统的技術中，分割图

像主要利用图像的边缘特征和阈值特征，而在深度学习的大势之下，图像分割开始有了突破性的发展。目前，深度学习主要运用在图像语义分割（Semantic Segmentation）中，即把图像分割成具有一定语义的块，如图 1.11 所示。

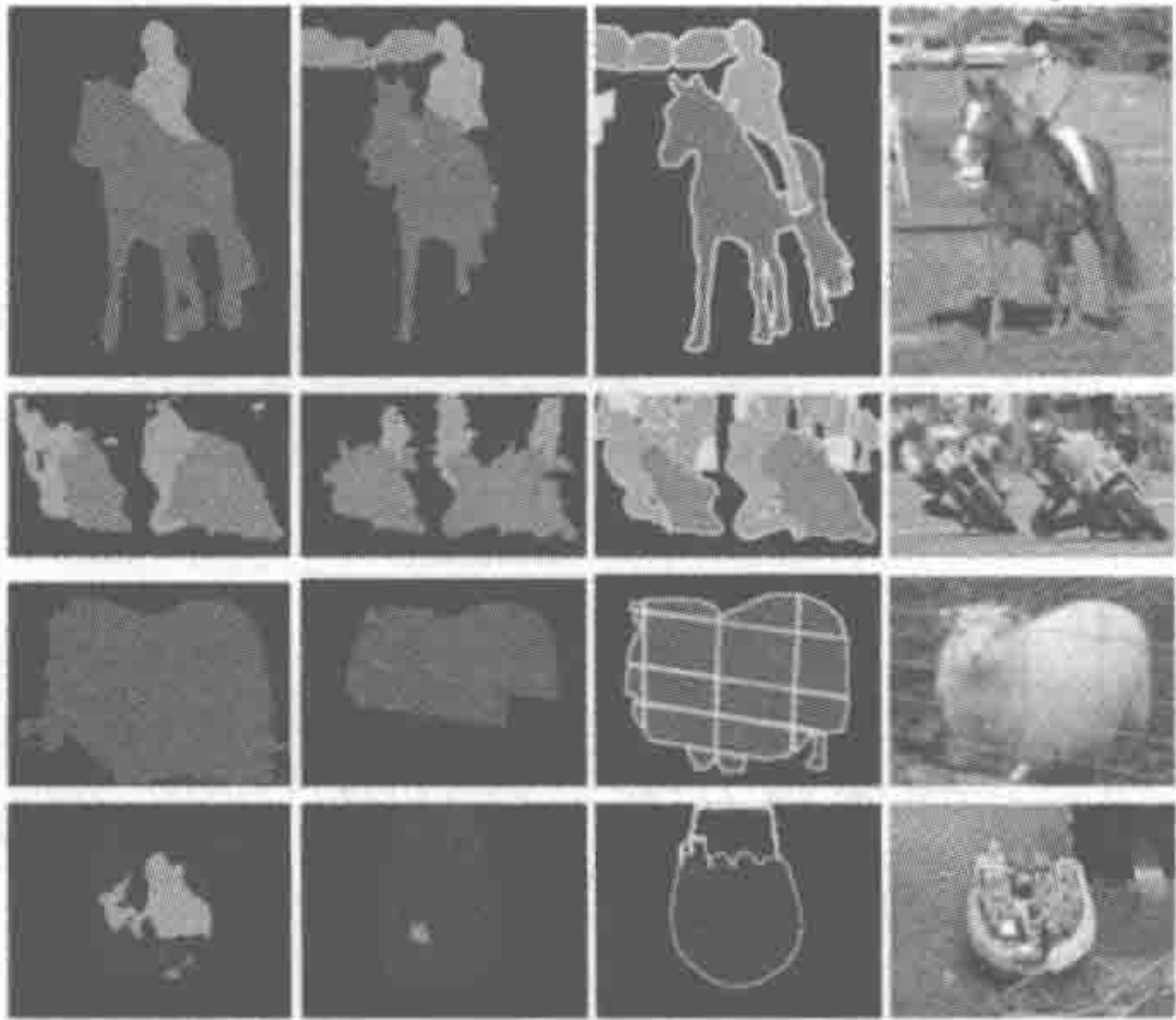


图 1.11

1.5.4 图像描述

图像描述（Image Caption）是通过深度学习的方法给输入的图像添加标题或者描述，如图 1.12 所示。

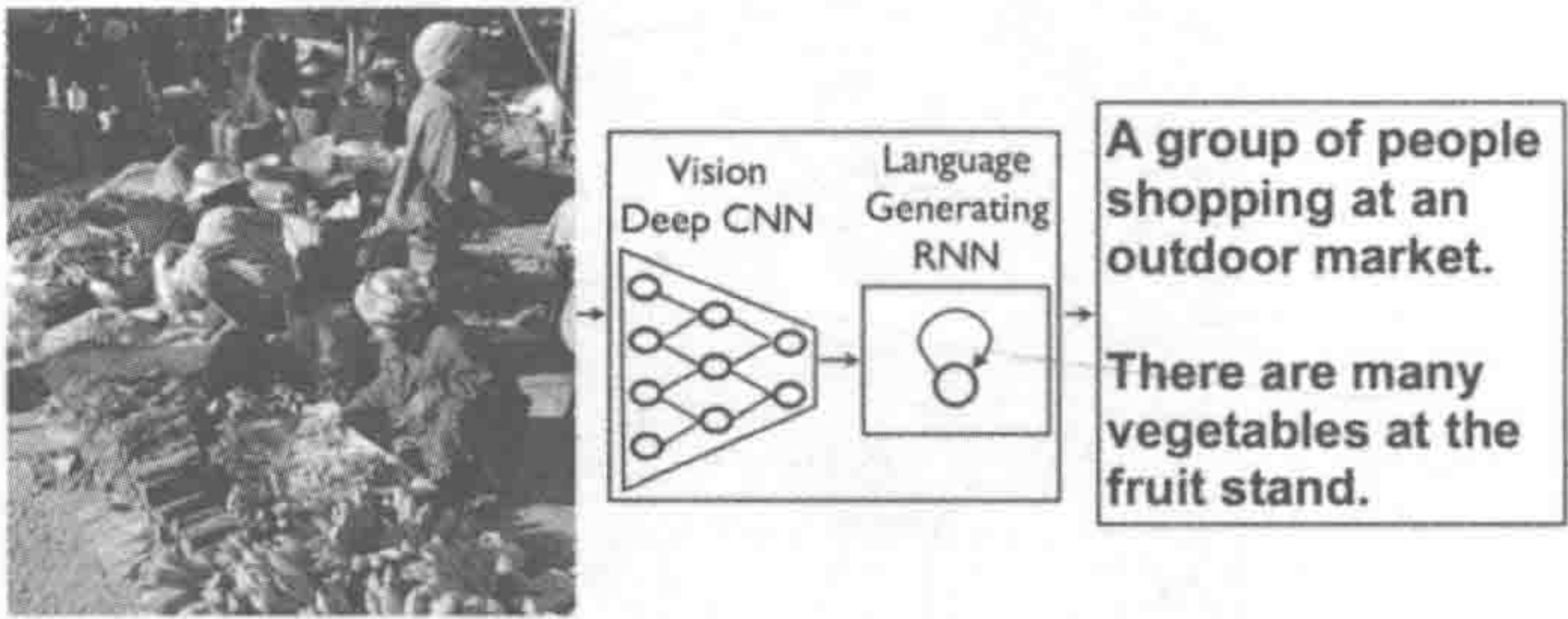


图 1.12

机器通过识别图像中的物体与场景，然后根据物体组织语言，最后形成一句描述：“A group of people shopping at an outdoor market”。图像描述包含了图像识别、图像检测和自然语言处理等，属于深度学习中的综合性研究。

1.5.5 机器翻译

机器翻译（Machine Translation）一直是人们不断探索的问题，世界上语言种类繁多，很