



计 算 机 科 学 丛 书

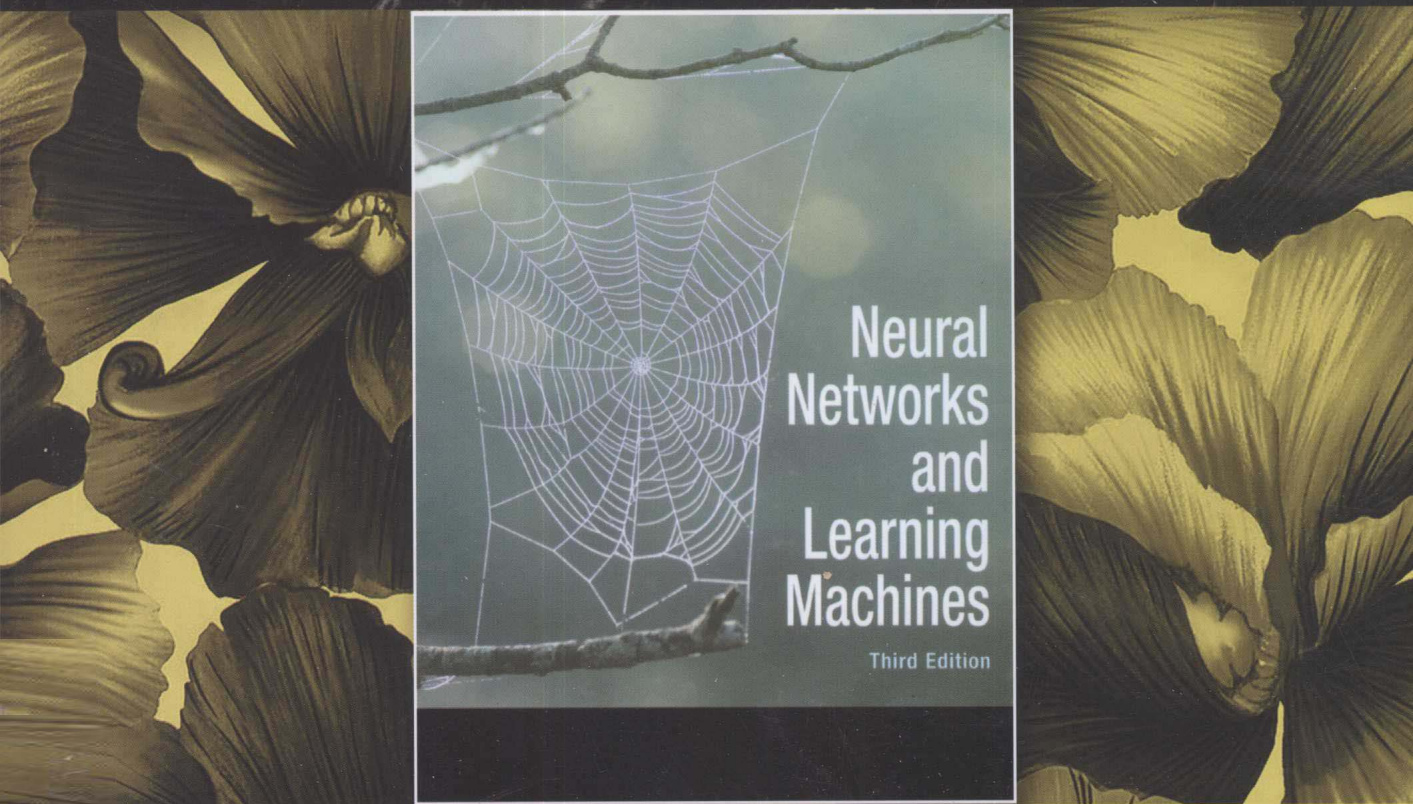
PEARSON

原书第3版

神经网络与机器学习

(加) Simon Haykin 著 申富饶 徐焯 郑俊 晁静 译

Neural Networks and Learning Machines
Third Edition



机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

原书第3版

神经网络与机器学习

(加) Simon Haykin 著 申富饶 徐焯 郑俊 晁静 译

Neural Networks and Learning Machines

Third Edition



机械工业出版社
China Machine Press

本书是关于神经网络的全面的、彻底的、可读性很强的、最新的论述。全书共 15 章，主要内容包括 Rosenblatt 感知器、通过回归建立模型、最小均方算法、多层感知器、核方法和径向基函数网络、支持向量机、正则化理论、主分量分析、自组织映射、信息论学习模型、动态规划、神经动力学、动态系统状态估计的贝叶斯滤波等。

本书适合作为高等院校计算机相关专业研究生及本科生的教材，也可供相关领域的工程技术人员参考。

Simplified Chinese edition copyright © 2011 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *Neural Networks and Learning Machines, Third Edition* (ISBN 978-0-13-147139-9) by Simon Haykin, Copyright © 2009.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

本书封面贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

封底无防伪标均为盗版

版权所有, 侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2009-1343

图书在版编目 (CIP) 数据

神经网络与机器学习 (原书第 3 版)/(加) 海金 (Haykin, S.) 著; 申富饶等译. —北京: 机械工业出版社, 2011. 1

(计算机科学丛书)

书名原文: *Neural Networks and Learning Machines, Third Edition*

ISBN 978-7-111-32413-3

I. 神… II. ①海… ②申… III. ①人工神经-神经网络 ②机器学习 IV. TP18

中国版本图书馆 CIP 数据核字 (2010) 第 213807 号

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: 迟振春

三河市明辉印装有限公司印刷

2011 年 3 月第 1 版第 1 次印刷

185mm×260mm·37.25 印张

标准书号: ISBN 978-7-111-32413-3

定价: 79.00 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) 88378991; 88361066

购书热线: (010) 68326294; 88379649; 68995259

投稿热线: (010) 88379604

读者信箱: hzsj@hzbook.com

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：www.hzbook.com

电子邮件：hzsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



从 20 世纪 40 年代 M-P 神经元模型的提出开始,神经网络的发展过程可谓是一波三折。1965 年 M. Minsky 和 S. Papert 的《感知机》使得神经网络的研究停滞了超过 10 年,直到 20 世纪 80 年代初 Hopfield 网络和误差反向传播算法等的提出,神经网络的研究才步入恢复期。时至今日,神经网络系统研究的重要意义已经得到广泛承认,在模式识别、人工智能、通信、控制、金融、机器人、生物信息学等许多领域都有广泛应用。可以说神经网络作为目前非线性科学和计算智能研究的主要内容之一,已经成为解决很多实际问题的一种必要的技术手段。

本书作者 Simon Haykin 长期从事神经网络的研究,其关于神经网络的系列教材是国际上最有影响力的教材之一。本书是其经典教材《Neural Networks: A Comprehensive Foundation》的第 3 版。正如本书的题目所示,这一版对神经网络和机器学习这两个密切相关的分支进行了全面分析,在前一版的基础上作了广泛修订,提供了神经网络和机器学习这两个重要性持续增长的学科的最新分析。本书全面、系统地介绍了神经网络的基本模型、基本方法,对神经网络的基本模型和主要学习理论作了深入研究,对神经网络的最新发展趋势和主要研究方向进行了全面而综合的介绍。

在翻译过程中,译者常常为本书作者严谨的治学态度及本书博大精深的内容而赞叹不已。本书综合了诸多神经网络和机器学习的最新研究,在翻译过程中虽然力求准确地反映原著内容,但由于译者水平有限,翻译中如有错漏之处,恳请读者批评指正。

本书的翻译得到了国家自然科学基金的资助(项目编号 60975047),特此表示感谢。同时,感谢参与本书翻译的全体人员,没有他们的辛勤工作,本书的中文译本是无法顺利完成的;感谢本书第 2 版(《神经网络原理》)的译者,在翻译过程中我们大量参考了第 2 版中文译本的内容;还要感谢南京大学计算机软件新技术国家重点实验室的支持。

南京大学计算机科学与技术系
计算机软件新技术国家重点实验室
申富饶
2010 年 10 月于南京

在写这本经典书籍第3版的时候，我遵循了本书第1版的基本原则：写一本关于神经网络的全面的、彻底的、可读性很强的、最新的论述。

新版本更名为《神经网络与机器学习》，主要是为了反映以下两个事实：

1. 感知器、多层感知器、自组织映射及神经动力学，以及其他一些通常被看成是神经网络一部分的主题，这些内容源自人类大脑所激发的灵感。

2. 核方法，例如支持向量机和核主分量分析，这些内容源自统计学习理论。

虽然这两者之间确实有一些共同的基本概念和应用，但是在神经网络和机器学习的具体操作存在一些微妙的差别。因而，如果将这两者放在同一个体系下共同研究，一些潜在的主题会变得更丰富，特别体现在以下方面：

- 将神经网络和机器学习的思想综合起来以完成更难的学习任务。这些学习任务往往是神经网络或者机器学习自身所无法解决的。
- 源自人类大脑的灵感往往会引起新的特别重要的新发现。

除此之外，本书的范围也有所扩大，提供了详细的动态规划和逐次状态估计，这两者各自都能够在一些重要方面影响强化学习和监督学习。

本书的组织

本书从导言部分开始，导言主要讲述了编写本书的动机，可作为后续章节的阅读基础。本书包括以下6个部分：

1. 第1~4章构成了本书的第一部分，主要介绍监督学习的一些经典方法。具体介绍如下：

- 第1章描述 Rosenblatt 感知器，重点介绍感知器收敛定理，以及在高斯环境下感知器和贝叶斯分类器的关系。
- 第2章讲述作为模型建立基础的最小二乘法，建立了在特定的高斯环境下这一方法和贝叶斯推理之间的关系。这一章还讨论了用于模式选择的最小描述长度（MDL）算法。
- 第3章讲述最小均方（LMS）算法及其收敛分析。其理论框架的分析揭示出两个原理：Kushner 直接法和朗之万（Langevin）方程（在非平衡态热力学中很著名）。

这三章通过对不同概念的介绍揭示了其共同特点：它们都是基于一个计算单元。更为重要的是，它们从各自的角度深入、细致地讨论了学习过程的深层知识——这一特征将在后续章节中进一步探讨。

第4章是关于多层感知器的，是 Rosenblatt 感知器的广义版本。这一相对比较长的章节包含如下主题：

- 反向传播算法、其优点和局限性，以及将其作为一个最优化方法来计算偏导数。
- 学习率的最优退火和自适应控制。
- 交叉验证。

- 卷积网络，来自于 Hubel 和 Wiesel 在视觉系统方面的开拓性研究。
- 将监督学习视为最优化问题，集中讨论共轭梯度法、拟牛顿法以及 Marquardt-Levenberg 算法。
- 非线性滤波。
- 最后，对于小规模和大规模学习问题作了对比。

2. 第二部分包括第 5 章和第 6 章，讨论了基于径向基函数 (RBF) 网络的核方法。

从某种意义上来说，第 5 章可以看做是对核方法的深入介绍。具体来说，这一章包括如下几个方面。

- 介绍 Cover 定理来作为对 RBF 网络的构造结构的理论证明。
- 描述相对简单的用于监督学习的两阶段混合过程，第一阶段基于聚类思想（即 K -均值算法）来计算隐藏层，第二阶段利用 LMS 或者最小二乘法来计算网络的线性输出层。
- 介绍核回归及其与 RBF 网络的关系。

第 6 章介绍支持向量机 (SVM)，通常这一方法被认为是一种监督学习方法。本质上 SVM 是一个两类分类器，本章中将包括如下几个主题：

- 定义在一对线性可分的两类之间最大分离边缘的条件。
- 当两个类是线性可分或者不可分时用来寻找最优超平面的二次最优化。
- 将 SVM 视为核机器，包含关于核欺骗和 Mercer 定理的讨论。
- SVM 的设计原理。
- ϵ -不敏感损失函数及其在回归问题最优化中的作用。
- 表示定理及希尔伯特空间构想和再生核希尔伯特空间构想 (RKHS) 的作用。

根据以上描述，很明显支持向量机的基本理论是建立在很强的数学背景之上的，因而 SVM 可以作为监督学习的一个具有强大计算能力的、一流的工具。

3. 本书第三部分只有一章——第 7 章。这一章介绍作为机器学习核心的正则化理论。本章将详细探讨如下几个主题：

- 建立在第 6 章讨论过的 RKHS 基础之上的 Tikhonov 经典正则化理论。这一理论隐含了一些深奥的数学概念：Tikhonov 泛函的 Fréchet 微分、Riesz 表示定理、Euler-Lagrange 方程、Green 函数，以及多变量高斯函数。
- 广义 RBF 网络及其计算精确性的修正。
- 正则最小二乘估计，根据表示定理的再讨论。
- 正则化参数估计，利用 Wahba 的广义交叉验证概念。
- 半监督学习，利用有标签和无标签样本。
- 可微流形及其在流形正则化中的作用——设计半监督学习机的基础。
- 寻找用于半监督学习的 RBF 网络中高斯核函数的光谱图理论。
- 处理半监督核机器的广义表示定理。
- 用于计算 RBF 网络线性输出层的拉普拉斯正则最小二乘 (LapRLS) 算法。这里需要说明的是，当内在正则化参数（对应于无标签数据）衰减为 0 的时候，算法相应地衰减为通常的最小二乘法。

这一高度理论化的章节具有非常实际的重要意义。首先，它提供了关于监督学习机的正则化基础。其次，它打下了设计正则化半监督学习机的基础。

4. 第 8~11 章构成本书的第四部分，讨论非监督学习。从第 8 章开始介绍由神经生物学研究直接激发的自组织的四个原则。

- 1) 自增强学习的 Hebb 假定。

- 2) 单个神经元或者一组神经元的突触连接为了有限的资源而进行的竞争。
- 3) 在胜利神经元及其邻居间的合作。
- 4) 包含于输入数据中的结构信息（如冗余）。

这一章的主要主题包括三个方面：

- 原则 1)、2) 和 4) 应用于单个神经元，最大特征滤波的 Oja 规则来源于这些原则；通过自组织获得的结果是值得注意的，它包含了自底向上和自顶向下学习。其次，最大特征滤波思想被推广到主分量分析 (PCA) 中，用来对输入数据进行维数削减，其所得算法称为广义 Hebb 算法 (GHA)。
- 本质上 PCA 是线性方法，因而其计算能力局限于二阶统计量。为了处理高阶统计量，核方法以类似于第 6 章支持向量机的相似方式应用于 PCA，但是和 SVM 的根本上的不同在于，核 PCA 是非监督方式。
- 遗憾的是，在处理自然图像的时候，核 PCA 从计算的角度变得很难操控。为了克服这一计算局限性，把 GHA 和核 PCA 结合起来组成一个新的在线非监督学习算法，称为核 Hebb 算法 (KHA)，这一方法可以用于图像去噪。

KHA 的产生是一个将机器学习的想法和来源于神经网络的补充想法结合起来的杰出例子，结合所产生的新算法克服了它们各自的实际局限性。

第 9 章介绍自组织映射 (SOM)，对自组织映射的开发遵从第 8 章介绍的自组织原则。从计算角度来说，自组织映射是一个简单的算法，而且具有内在的构造拓扑映射的强大能力，它包括如下一些有用的特性：

- 从空间上离散逼近输入空间，负责数据生成。
- 拓扑次序，在某种意义上神经元的空间位置在拓扑图上对应于输入空间中的特定特征。
- 输入输出密度匹配。
- 输入数据特征选择。

SOM 在实际中被广泛应用，构造上下文映射和分层次矢量量化被作为 SOM 运算能力的两个有说服力的例子。事实上，令人惊异的是，尽管 SOM 展示了多个有趣的特性并且能够解决很难的计算任务，但它依然缺少一个能用来最优化的目标函数。为了填补这一缺口，以提供改进拓扑映射的可能性，自组织映射采用了核方法。这一改进是通过引入一个熵函数作为目标函数并且最大化这个函数来实现的。我们再次看到了将来自于神经网络的思想 and 补充的核理论思想结合所带来的实际好处。

第 10 章探讨如何将来自于香农 (Shannon) 信息论的原则作为工具来实现非监督学习。这一个相对较长的章节从回顾香农信息论开始，重点讨论了熵、互信息、相对熵 (KLD) 等概念。这一回顾也包括系词 (copula) 的概念，遗憾的是这一概念几十年来没有被注意到。更重要的是，系词提供了对一对相关随机变量之间统计相关性的测量。在任何事件中，集中于将互信息作为目标函数，这一章建立了如下原则：

- 最大互信息原则，最大化神经系统的输入和输出之间的互信息；最大互信息和冗余减少之间有着很紧密的关系。
- I_{max} 原则，最大化由相关输入驱动的神经系统对的单一输出之间的互信息。
- I_{min} 原则，以一种和 I_{max} 原则相似的方式操作，但这里是最小化输出随机变量对之间的互信息。
- 独立分量分析 (ICA) 原则，提供一种很强的工具用于盲分离来自统计独立源信号的隐藏集合。当满足一定的操作条件时，ICA 原则将提供对源信号进行恢复的起源程序基础，用于恢复的信号来自于对源信号的线性混合变形的相应的观察集合。这里将介

绍两个特别的 ICA 算法。

- 1) 自然梯度学习算法，除了拉伸和排列之外，通过最小化参数概率密度函数和相应的阶乘分布之间的 KLD 来解决 ICA 问题。
- 2) 最大熵学习算法，最大化反混合输出的非线性变换版本的熵；这一算法通常被认为是 ICA 的最大化信息算法，也表现出拉伸和排列性质。

第 10 章还描述了另一个称为快速 ICA (FastICA) 的重要的 ICA 算法，这一算法正如其名字那样，计算速度快。这一算法基于负熵的概念最大化对比函数，对比函数提供了对于随机变量的非高斯分布程度的测量。作为 ICA 的延续，本章继续描述了一种称为相关 ICA 的新算法，其开发是根据最大化信息和 Imax 原则的融合并经由连接函数的运用来完成的；相关 ICA 在采集调幅信号的混合物的包迹时非常有用。最后，第 10 章介绍了另一个来自于香农信息论的称为速率失真理论的概念，这一理论被用来开发这一章的最后一个概念：信息瓶颈。给定关于输入向量和（有关的）输出向量的连接分布，这一方法通过如下方式被构造为约束最优化问题：在两个信息量之间做一个权衡，一个信息量是关于输入的瓶颈向量中包含的信息，另一个信息量是关于输出的瓶颈向量中所包含的信息。这一章将利用信息瓶颈法来寻找数据表达的最优流形。

第 11 章讲述非监督学习的最后途径，利用源自统计力学的随机方法来实现。统计力学的研究和信息论密切相关。这一章从回顾 Helmholtz 自由能和熵概念（从统计力学意义上）开始，紧接着介绍马尔可夫链。然后介绍用于产生马尔可夫链的 Metropolis 算法，其转移概率将收敛到唯一的、稳定的分布。接下来以两个方面作为随机方法讨论的结束：一是用于全局最优化的模拟退火，二是 Gibbs 抽样，它可以作为 Metropolis 算法的特殊形式。有了手头这些统计力学的背景知识，就可以讲述 Boltzmann 机了，Boltzmann 机从历史上来说是文献中讨论的第一个多层学习机器。遗憾的是，Boltzmann 机的学习过程非常慢，特别是当隐藏神经元的数目很大的时候，因而其实用性是最主要的缺陷。人们提出了很多变种方法来克服 Boltzmann 机的缺点。其中到目前为止最成功的创新方法是深度信念网络，它明智地把下面的两个功能组合起来形成了一个高效的机器：

- 生成模型，无监督地一层一层自底向上学习所得结果。
- 推论，自顶向下学习所得结果。

最后，第 11 章讲述确定性退火来克服模拟退火极端的计算需求问题；确定性退火的问题在于其可能陷入局部极小点。

5. 到目前为止，本书集中精力讲述了构造用于监督学习、半监督学习和非监督学习的算法。第 12 章，作为本书下一个部分，是关于强化学习的。强化学习以一种在线方式发生，作为智能体（如机器人）与其周围的环境相互作用的结果。实际上，动态规划是强化学习的核心。相应地，第 15 章的前面部分用来介绍 Bellman 动态规划方法，然后用来证明两个广泛使用的强化学习方法：时序差分学习 (TD) 和 Q-学习，这两种方法能通过作为动态规划的特例推导得出。TD 学习和 Q-学习都是相对比较简单在线强化学习算法，无需转移概率知识。然而，其实际应用局限于状态空间的维数处于中等程度的情况。在大规模动态系统中，维数灾难变得非常严重，使得不仅仅是动态规划，也包括其近似形式的 TD 学习和 Q-学习变得难以计算。为了克服这一严重的局限性，这一章描述了两个逼近动态规划的非直接方法：

- 线性方法，称为最小二乘策略评估 (LSPV) 算法。
- 非线性方法，利用神经网络（如多层感知器）作为通用逼近器。

6. 本书最后一部分包括第 13、14 和 15 章，讨论非线性反馈系统，特别强调递归神经网络：

1) 第 13 章研究神经动力学，对稳定性问题给予了特别的关注。这一章介绍了 Lyapunov 直接法，这个方法包含两个定理，一个用来处理系统稳定性，另一个用来处理渐近稳定性。这一方法的核心是 Lyapunov 函数，通常来说能量函数就能满足这一函数的要求。有了这样的背景知识，就可以引出两种联想记忆模型：

- Hopfield 模型，这一模型的操作说明一个复杂的系统是能够产生简单的突现行为的。
- 盒中脑状态模型，它是聚类的基础。

第 13 章还讨论了混沌过程的特性及其动态重构的正则化过程。

2) 第 14 章是关于贝叶斯滤波器的，贝叶斯滤波器至少从概念意义上提供了逐次状态估计算法的统一基础。这一章的发现总结为以下几点：

- 经典的线性高斯环境下的卡尔曼滤波器可以通过利用最小均方差准则来推导；在这一章最后的一个习题中，证明这样推导的卡尔曼滤波器是贝叶斯滤波器的特例。
- 平方根滤波用来克服卡尔曼滤波在实际应用中遇到的发散现象。
- 扩展卡尔曼滤波（EKF）用来解决动力系统中非线性属于软排序的情况；保持高斯假设。
- 以一个新的称为数值积分卡尔曼滤波器（CKF）的滤波器为例来证明贝叶斯滤波器的直接逼近形式。这里再次强调了保持高斯假设。
- 以粒子滤波器为例来证明贝叶斯滤波器的非直接逼近形式，粒子滤波器的实现能够调节非线性程度和非高斯程度。

卡尔曼滤波本质上是预测-改正机制，第 14 章接着描述“类卡尔曼滤波”在人类大脑的一定区域的可能作用。

本书第 15 章研究动态驱动的递归神经网络。这一章的开始部分讨论不同的递归网络结构（模型）及其计算能力，紧接着介绍训练递归网络的两个算法：通过时间的反向传播和实时递归学习。

遗憾的是，这两个方法都是基于梯度的，容易遭遇所谓的消失梯度（vanishing-gradient）问题。为减轻这一问题，本书较详细地讨论了利用非线性逐次状态估计，采用全新的方式来对递归网络进行监督训练。这里，对于扩展卡尔曼滤波器（简单，但是导数依赖）以及数值积分卡尔曼滤波器（导数自由，但是数学上更加复杂）作为监督学习的逐次状态估计器的优缺点进行了讨论。此外，还讨论了对递归网络来说唯一的自适应行为的出现以及利用自适应技巧来增强递归网络性能的潜在好处。

在本书不同部分出现的一个重要的主题是，将监督学习和半监督学习应用于大规模问题。这包括本书评论中所指出的这一主题还处于发展的初期阶段；更重要的是，本书还为这一问题的未来发展描述了四阶段过程。

本书特色

本书完整、详尽地讨论了各个主题，除此之外，本书还有以下几个截然不同的特色：

1. 第 1~7 章以及第 10 章包含计算机实验，涉及双月形态，为两类分类问题产生数据。实验涵盖了从简单的线性可分模式例子到困难的不可分模式例子。作为运行例子的双月形态，被用于第 1~7 章以及第 10 章，因而提供了一个用于研究和比较这 8 章中描述的算法的实验途径。

2. 针对第 8 章的主分量分析、第 9 章的 SOM 和核 SOM，以及第 15 章的利用 EKF 和 CKF 算法对 Mackay-Glass 吸引子进行动态重构等，也进行了计算机实验。

3. 给出了几个利用现实数据进行研究的例子：

- 第 7 章讨论了利用拉普拉斯 RLS 算法对美国邮政服务 (USPS) 数据进行半监督学习。
- 第 8 章讨论了如何将 PCA 应用于手写数字数据, 并描述了如何对图像进行编码和去噪。
- 第 10 章利用稀疏传感编码和 ICA 对自然图像进行分析。
- 第 13 章利用正则 RBF 网络将动态重构应用于 Lorenz 吸引子。

第 15 章也包含了一节关于模型参照自适应控制系统的案例研究。

4. 每一章的最后都有注释和参考文献用于进一步学习, 每章末尾还提供了习题, 用来练习并丰富读者的专业知识。

本书的“术语”表也进行了扩充, 包含了用于处理矩阵分析和概率论问题的方法学解释。

5. 本书所有图和表格的 PowerPoint 文件都可以提供给教师, 可到华章网站 (www.hzbook.com) 下载。

我们尽了最大努力来使本书不犯错误, 更重要的是, 我们也尽力提高它的可读性。

Simon Haykin
于 Ancaster, Ontario

缩写

AR	autoregressive	自回归
BBTT	back propagation through time	通过时间的反向传播
BM	Boltzmann machine	Boltzmann 机
BP	back propagation	反向传播
b/s	bits per second	每秒比特率
BSB	brain-state-in-a-box	盒中脑状态
BSS	Blind source (signal) separation	盲源 (信号) 分离
cmm	correlation matrix memory	相关矩阵记忆
CV	cross-validation	交叉验证
DFA	deterministic finite-state automata	确定性有限状态自动机
EKF	extended Kalman filter	扩展卡尔曼滤波器
EM	expectation-maximization	期望最大化
FIR	finite-duration impulse response	有限时间冲击响应
FM	frequency-modulated (signal)	频率调制 (信号)
GCV	generalized cross-validation	广义交叉验证
GHA	generalized Hebbian algorithm	广义 Hebb 算法
GSLC	generalized sidelobe canceler	广义旁瓣消除器
Hz	hertz	赫兹
ICA	independent-components analysis	独立分量分析
Infomax	maximum mutual information	最大互信息
Imax	variant of Infomax	最大互信息的变体
Imin	another variant of Infomax	最大互信息的另一个变体
KSOM	kernel self-organizing map	核自组织映射
KHA	kernel Hebbian algorithm	核 Hebb 算法
LMS	least-mean-square	最小均方
LR	likelihood ratio	似然比
LS	Least-squares	最小二乘

LS-TD	Least-squares, temporal-difference	最小二乘, 时序差分
LTP	long-term potentiation	长期增强
LTD	long-term depression	长期衰减
LRT	Likelihood ratio test	似然比测试
MAP	Maximum a posteriori	最大后验估计
MCA	minor-components analysis	次分量分析
MCMC	Markov Chain Monte Carlo	马尔可夫链蒙特卡罗
MDL	minimum description length	最小描述长度
MIMO	multiple input-multiple output	多输入多输出
ML	maximum likelihood	最大似然
MLP	multilayer perceptron	多层感知器
MRC	model reference control	模型参考控制
NARMA	nonlinear autoregressive moving average	非线性自回归滑动平均
NARX	nonlinear autoregressive with exogenous inputs	具有外部输入的非线性自回归
NDP	neuro-dynamic programming	神经动态规划
NW	Nadaraya-Watson (estimator)	Nadaraya-Watson (估计器)
NWKR	Nadaraya-Watson kernel regression	Nadaraya-Watson 核回归
OBD	optimal brain damage	最优脑损伤
OBS	optimal brain surgeon	最优脑外科
OCR	optical character recognition	光学字符识别
PAC	probably approximately correct	可能近似正确
PCA	principal-components analysis	主分量分析
PF	Particle Filter	粒子滤波器
pdf	probability density function	概率密度函数
pmf	probability mass function	概率质量函数
QP	quadratic programming	二次规划
RBF	radial basis function	径向基函数
RLS	recursive least-squares	递归最小二乘
RLS	regularized least-squares	正则最小二乘
RMLP	recurrent multilayer perceptron	递归多层感知器
RTRL	real-time recurrent learning	实时递归学习
SIMO	single input-multiple output	单输入多输出
SIR	sequential importance resampling	逐次重要重采样
SIS	sequential important sampling	逐次重要采样
SISO	single input-single output	单输入单输出
SNR	signal-to-noise ratio	信噪比
SOM	self-organizing map	自组织映射
SRN	simple recurrent network (also referred to as Elman's recurrent network)	简单递归网络 (也称为 Elman 递归网络)

SVD	singular value decomposition	奇异值分解
SVM	support vector machine	支持向量机
TD	temporal difference	时序差分
TDNN	time-delay neural network	时延神经网络
TLFN	time-lagged feedforward network	时间滞后前馈网络
VC	Vapnik-Chervononkis (dimension)	Vapnik-Chervononkis (维数)
VLSI	very-large-scale integration	超大规模集成
XOR	exclusive OR	异或

重要的符号

a	action	动作
$\mathbf{a}^T \mathbf{b}$	inner product of vectors \mathbf{a} and \mathbf{b}	向量 \mathbf{a} 和 \mathbf{b} 的内积
\mathbf{ab}^T	outer product of vectors \mathbf{a} and \mathbf{b}	向量 \mathbf{a} 和 \mathbf{b} 的外积
$\binom{l}{m}$	binomial coefficient	二项式系数
$A \cup B$	unions of A and B	A 和 B 的并集
B	inverse of temperature	温度的逆
b_k	bias applied to neuron k	神经元 k 的偏置
$\cos(\mathbf{a}, \mathbf{b})$	cosine of the angle between vectors \mathbf{a} and \mathbf{b}	向量 \mathbf{a} 和 \mathbf{b} 夹角的余弦
$c_{u,v}(u, v)$	probability density function of copula	系词的概率密度函数
D	depth of memory	记忆深度
$D_{f \parallel g}$	Kullback-Leibler divergence between probability density functions f and g	概率密度函数 f 和 g 之间的 Kullback-Leibler 散度
$\tilde{\mathbf{D}}$	adjoint of operator \mathbf{D}	算子 \mathbf{D} 的伴随矩阵
E	energy function	能量函数
E_i	energy of state i in statistical mechanics	统计力学中状态 i 的能量
\mathbb{E}	statistical expectation operator	统计期望算子
$\langle E \rangle$	average energy	平均能量
exp	exponential	指数
\mathcal{E}_{av}	average squared error, or sum of squared errors	平均平方误差或平方误差和
$\mathcal{E}(n)$	instantaneous value of the sum of squared errors	平方误差和的瞬时值
\mathcal{E}_{total}	total sum of error squares	总平方误差和
F	free energy	自由能量
\mathcal{F}^*	subset (network) with minimum empirical risk	经验风险最小的子集 (网络)
\mathbf{H}	Hessian (matrix)	Hessian 矩阵
\mathbf{H}^{-1}	inverse of Hessian \mathbf{H}	Hessian 矩阵 \mathbf{H} 的逆
i	square root of -1 , also denoted by j	-1 的平方根, 亦记作 j
\mathbf{I}	identity matrix	单位矩阵
\mathbf{I}	Fisher's information matrix	Fisher 信息矩阵
J	mean-square error	均方误差

J	Jacobian (matrix)	Jacobi 矩阵
$\mathbf{P}^{1/2}$	square root of matrix P	矩阵 P 的方根
$\mathbf{P}^{T/2}$	transpose of square root of matrix P	矩阵 P 的方根的转置
$\mathbf{P}_{n,n-1}$	error covariance matrix in Kalman filter theory	卡尔曼滤波理论中的误差协方差矩阵
k_B	Boltzmann constant	Boltzmann 常数
log	logarithm	对数
$L(\mathbf{w})$	log-likelihood function of weight vector w	权值向量 w 的对数似然函数
$\mathcal{L}(\mathbf{w})$	log-likelihood function of weight vector w based on a single example	单样本的权值向量 w 的对数似然函数
\mathbf{M}_c	controllability matrix	可控矩阵
\mathbf{M}_o	observability matrix	可观察矩阵
n	discrete time	离散时间
p_i	probability of state i in statistical mechanics	统计力学中状态 i 的概率
p_{ij}	transition probability from state i to state j	从状态 i 到状态 j 的转移概率
P	stochastic matrix	随机矩阵
$P(e \mathcal{C})$	conditional probability of error e given that the input is drawn from class \mathcal{C}	从类 \mathcal{C} 中输入时误差 e 的条件概率
P_α^+	probability that the visible neurons of a Boltzmann machine are in state α , given that the network is in its clamped condition (i. e. , positive phase)	假设网络处于钳制条件 (即正向阶段) 时, Boltzmann 机的可见神经元状态 为 α 的概率
P_α^-	probability that the visible neurons of a Boltzmann machine are in state α , given that the network is in its free-running condition (i. e. , negative phase)	假设网络处于自由运行条件 (即负向 阶段) 时, Boltzmann 机的可见神经 元状态为 α 的概率
$\hat{r}_x(j, k; n)$	estimate of autocorrelation function of $x_j(n)$ and $x_k(n)$	$x_j(n)$ 和 $x_k(n)$ 的自相关函数估计
$\hat{r}_{dx}(k; n)$	estimate of cross-correlation function of $d(n)$ and $x_k(n)$	$d(n)$ 和 $x_k(n)$ 的交叉相关函数估计
R	correlation matrix of an input vector	输入向量的相关矩阵
t	continuous time	连续时间
T	temperature	温度
\mathcal{T}	training set (sample)	训练集 (样本)
tr	operator denoting the trace of a matrix	表示矩阵迹的算子
var	variance operator	方差算子
$V(\mathbf{x})$	Lyapunov function of state vector x	状态向量 x 的 Lyapunov 函数
v_j	induced local field or activation potential of neuron j	神经元 j 的诱导局部域或激活位势
\mathbf{w}_o	optimum value of synaptic weight vector	突触权值向量的最优值
w_{kj}	weight of synapse j belonging to neuron k	属于神经元 k 的突触 j 的突触权值
\mathbf{w}^*	optimum weight vector	最优权值向量
$\bar{\mathbf{x}}$	equilibrium value of state vector x	状态向量 x 的平衡值
$\langle x_j \rangle$	average of state x_j in a "thermal" sense	"热" 意义下状态 x_j 的平均

\hat{x}	estimate of x , signified by the use of a caret (hat)	x 的估计, 用加字符号 $\hat{\quad}$ (帽符号) 表示
$ x $	absolute value (magnitude) of x	x 的绝对值 (幅度)
x^*	complex conjugate of x , signified by asterisk as superscript	状态 x 的复共轭, 用星号 $*$ 作上标
$\ \mathbf{x}\ $	Euclidean norm (length) of vector \mathbf{x}	向量 \mathbf{x} 的欧几里得范数 (长度)
\mathbf{x}^T	transpose of vector \mathbf{x} , signified by the superscript T	向量 \mathbf{x} 的转置, 用上标 T 表示
z^{-1}	unit-time delay operator	单位时间延迟算子
Z	partition function	剖分函数
$\delta_j(n)$	local gradient of neuron j at time n	神经元 j 在时刻 n 的局部梯度
Δw	small change applied to weight w	权值 w 的微小改变
∇	gradient operator	梯度算子
∇^2	Laplacian operator	拉普拉斯算子
$\nabla_w J$	gradient of J with respect to w	J 关于 w 的梯度
$\nabla \cdot \mathbf{F}$	divergence of vector \mathbf{F}	向量 \mathbf{F} 的散度
η	learning-rate parameter	学习率参数
κ	cumulant	累积量
μ	policy	策略
θ_k	threshold applied to neuron k (i. e., negative of bias b_k)	神经元 k 的阈值 (即偏置 b_k 的负值)
λ	regularization parameter	正则化参数
λ_k	k th eigenvalue of a square matrix	方阵的第 k 个特征值
$\varphi_k(\cdot)$	nonlinear activation function of neuron k	神经元 k 的非线性激活函数
\in	symbol for “belongs to”	“属于”符号
\cup	symbol for “union of”	“并”符号
\cap	symbol for “intersection of”	“交”符号
$*$	symbol for convolution	卷积符号
$+$	superscript symbol for pseudoinverse of a matrix	矩阵伪逆的上标符号
$+$	superscript symbol for updated estimate	更新估计的上标符号

开区间和闭区间

- 变量 x 的开区间 (a, b) 表示 $a < x < b$ 。
- 变量 x 的闭区间 $[a, b]$ 表示 $a \leq x \leq b$ 。
- 变量 x 的半闭半开区间 $[a, b)$ 表示 $a \leq x < b$; 类似地, 变量 x 的半开半闭区间 $(a, b]$ 表示 $a < x \leq b$ 。

最小和最大

- 符号 $\arg \min_{\mathbf{w}} f(\mathbf{w})$ 表示函数 $f(\mathbf{w})$ 关于变元向量 \mathbf{w} 的最小值。
- 符号 $\arg \max_{\mathbf{w}} f(\mathbf{w})$ 表示函数 $f(\mathbf{w})$ 关于变元向量 \mathbf{w} 的最大值。

记号 I：矩阵分析

标量：用小写斜体符号表示标量。

向量：用小写粗体符号表示向量。

向量被定义为一列标量。因而 m 维向量 \mathbf{x} 和 \mathbf{y} 的内积可以写成

$$\mathbf{x}^T \mathbf{y} = [x_1, x_2, \dots, x_m] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \sum_{i=1}^m x_i y_i$$

其中，上标 T 用来表示矩阵转置。其内积为标量，因而我们有

$$\mathbf{y}^T \mathbf{x} = \mathbf{x}^T \mathbf{y}$$

矩阵：用大写粗体符号表示矩阵。

矩阵相乘是通过行和列的相乘来计算的。为了说明这一点，考虑 $m \times k$ 的矩阵 \mathbf{X} 和 $k \times l$ 的矩阵 \mathbf{Y} 。这两个矩阵的乘积产生一个 $m \times l$ 的矩阵

$$\mathbf{Z} = \mathbf{XY}$$

更具体地说，矩阵 \mathbf{Z} 的第 ij 个分量是通过矩阵 \mathbf{X} 的第 i 行和矩阵 \mathbf{Y} 的第 j 列相乘而得到的，这两者都由 k 个标量组成。

一对 m 维向量 \mathbf{x} 和 \mathbf{y} 的外积写成 \mathbf{xy}^T ，是一个 $m \times m$ 的矩阵。

记号 II：概率论

随机变量：用大写的斜体符号来表示随机变量。随机变量的样本值（即单次实现）用相应的小写斜体符号来表示。例如，我们用 X 来表示随机向量，而用 x 来表示其样本值。

随机向量：用大写的粗体符号来表示随机向量。相似地，随机向量的样本值用相应的小写粗体符号来表示。例如，我们用 \mathbf{X} 来表示随机向量，而用 \mathbf{x} 来表示其样本值。

随机变量 X 的概率密度函数 (pdf) 由 $p_x(\mathbf{x})$ 来表示，这是关于样本值 \mathbf{x} 的函数；其下标 \mathbf{X} 是用来提示 pdf 是关于随机向量 \mathbf{X} 的。