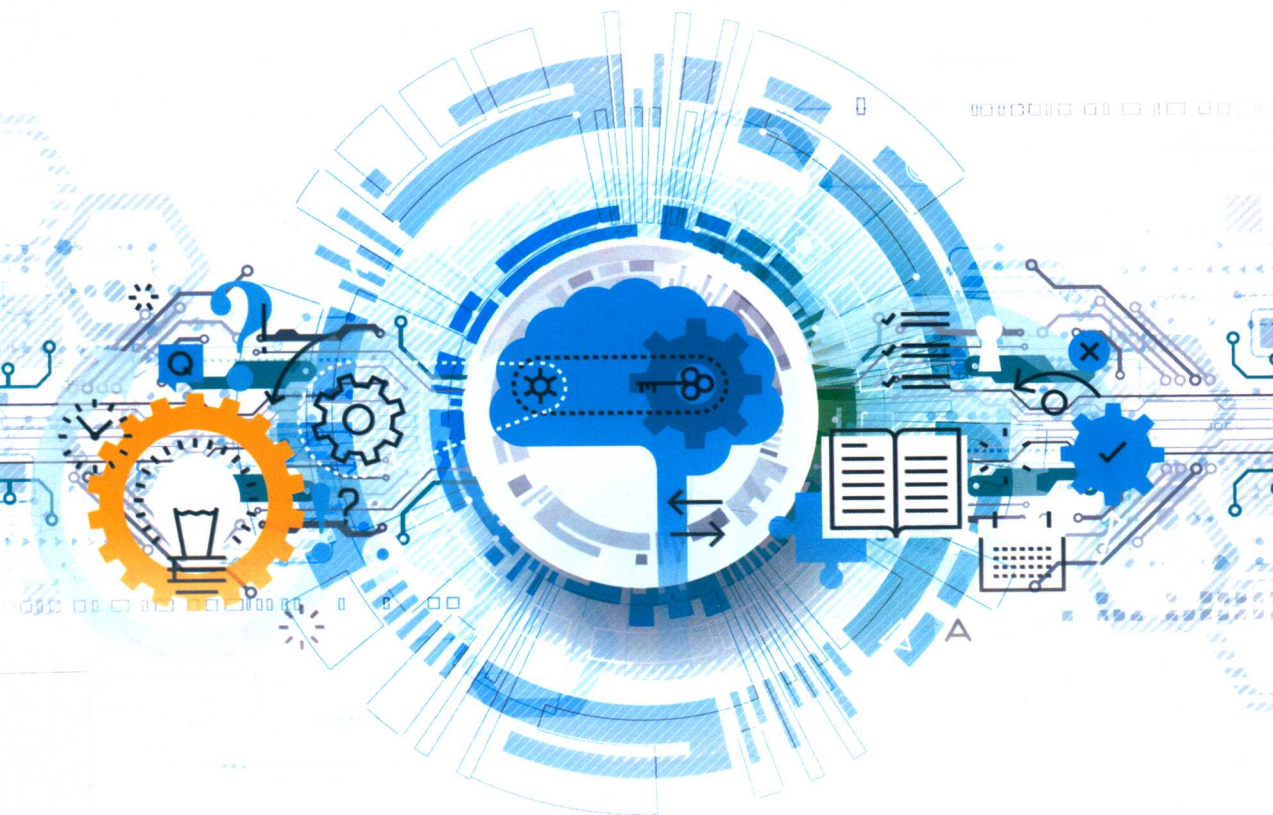


MATLAB Machine Learning

MATLAB与机器学习

[美] 迈克尔·帕拉斯泽克 (Michael Paluszek) 著
斯蒂芬妮·托马斯 (Stephanie Thomas)
李三平 陈建平 译



机械工业出版社
China Machine Press

智能系统与技术丛书

智能系统与技术丛书

MATLAB 与机器学习 (英文) 迈克尔·帕拉斯泽克 (Michael Paluszek) 著
斯蒂芬妮·托马斯 (Stephanie Thomas) 著 李三平 陈建平 译
机械工业出版社

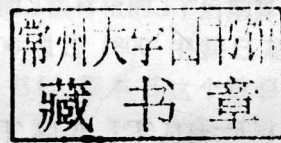
ISBN 978-7-111-28984-6

1. M... ① 李三平 陈建平 译 ② 李三平 陈建平 译
IV. TP181

MATLAB Machine Learning

MATLAB与机器学习

[美] 迈克尔·帕拉斯泽克 (Michael Paluszek) 著
斯蒂芬妮·托马斯 (Stephanie Thomas) 著
李三平 陈建平 译



图书在版编目 (CIP) 数据

MATLAB 与机器学习 / (美) 迈克尔·帕拉斯泽克 (Michael Paluszek), (美) 斯蒂芬妮·托马斯 (Stephanie Thomas) 著; 李三平, 陈建平译. —北京: 机械工业出版社, 2018.1
(智能系统与技术丛书)

书名原文: MATLAB Machine Learning

ISBN 978-7-111-58984-6

I. M… II. ① 迈… ② 斯… ③ 李… ④ 陈… III. Matlab 软件—应用—机器学习
IV. TP181

中国版本图书馆 CIP 数据核字 (2018) 第 011589 号

本书版权登记号: 图字 01-2017-7340

Michael Paluszek, Stephanie Thomas: MATLAB Machine Learning (ISBN: 978-1-4842-2249-2).
Original English language edition published by Apress Media.

Copyright © 2017 by Michael Paluszek, Stephanie Thomas. Simplified Chinese-language edition
copyright © 2018 by China Machine Press. All rights reserved.

This edition is licensed for distribution and sale in the People's Republic of China only, excluding
Hong Kong, Taiwan and Macao and may not be distributed and sold elsewhere.

本书原版由 Apress 出版社出版。

本书简体字中文版由 Apress 出版社授权机械工业出版社独家出版。未经出版者预先书面许可, 不得以任何方式
复制或抄袭本书的任何部分。

此版本仅限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) 销售发行, 未经授权的本书出口
将被视为违反版权法的行为。

MATLAB 与机器学习

出版发行: 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: 谢晓芳

责任校对: 李秋荣

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2018 年 2 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 19

书 号: ISBN 978-7-111-58984-6

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) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

FOREWORD

推荐序

对于很多读者来说，对 MATLAB 的利用可能还停留在单纯的仿真甚至矩阵运算方面。经过 30 多年的发展，MATLAB 积累了大量工业级的工具箱，广度上涵盖通用科学计算、通信与信号处理、控制系统、金融等各个领域。在深度上，采用基于模型的设计方法，MATLAB 已经在算法开发、系统设计、自动代码生成以及从单元测试到系统验证的各个方面，具备了成熟的流程和完整的功能。

在过去 30 年中，MATLAB 一直活跃在数据分析领域。大量的用户使用 MATLAB 进行数据分析，以获取数据的特征 (Data Analysis)，并对未知输入进行预测 (Data Analytics)。机器学习正是目前使用最为广泛的算法手段之一。

机器学习是一个系统工程，一个完整的数据分析流程包括数据的获取、数据清洗和探索、数据分析以及结果发布。这也是 MATLAB 作为统一开发环境的价值。在数据获取方面，MATLAB 支持多种类型的数据输入，包括分布式文件系统、硬件设备、测试仪器、数据库等，足以应付大多数的机器学习场景。同时，不同领域的工具箱能够提供专业的数据预处理和可视化功能。随着技术的进步，数据的总量正在以指数的速度增长，MATLAB 支持并行计算和云计算，能大大提升算法的研发效率。最后，MATLAB 强大的嵌入式代码生成和发布能力，提供了算法结果的一键式发布能力，极大减少了后期算法移植的工作量。

繁荣的图书市场上，在 MATLAB 图书计划中注册的书籍已超过 1800 部，涉及 MATLAB 的方方面面。虽然在售的 MATLAB 图书远远超过这个数目，但本书仍是我们 MATLAB 图书计划里一百多部中文书籍中第一本关于机器学习的专著。本书由 MathWorks 中国技术专家陈建平和 EMC 研究院李三平博士翻译。

从自主系统的角度，本书对机器学习的原理进行了介绍。我们在接触大量 MATLAB 用户的过程中，发现领域专家想利用机器学习的手段对搜集的数据进行分析和预测，但不知如何快速开展工作。也有不少人通过网络了解到一些机器学习的算法描述，却苦于无法找到兼顾原理与工具的中文图书，在有效地选取最佳算法与进行分析和预测等实践方面存在困难。本书为这些读者提供了重要的指南，有助于他们快速开展工作，有效选取合适的算法，并进行分析和预测。

我相信本书在 MATLAB 与机器学习方面系统性的阐述能够对读者有所帮助。

周拥华

MathWorks 中国技术经理

图书在版编目 (CIP) 数据

MATLAB 与机器学习 (第 4 版) 迈克尔·帕拉斯泽克 (Michael Paluszak) (美) 斯蒂芬·托马斯 (Stephane Thomas) 著; 李三平、陈建宇译. —北京: 机械工业出版社, 2019

(智能系统与技术丛书)

书名原文: MATLAB Machine Learning

ISBN 978-7-111-62907-3 THE TRANSLATORS' WORDS

译者序

机器学习 (Machine Learning) 正在又一次成为人工智能 (Artificial Intelligence) 发展热潮中的焦点技术, 它包括处理诸如图片、视频、语音、文字等感知数据的深度神经网络 (Deep Neural Networks, DNN) 技术, 属于人工智能项目中基础和难点技术的自动控制 and 自动驾驶 (Self-Driving) 等。这些领域正在不断取得令人激动的技术成就。

MATLAB 是由美国 MathWorks 公司出品的商业数学软件, 提供优秀的数值计算能力, 并且利用工具箱 (Toolbox) 来增强对各种工业领域的适用性。本书则详细介绍了 MATLAB 对机器学习领域中关键技术的实现与应用, 内容涵盖数据分类、面部识别、系统状态估计、自适应控制和自动驾驶等经典学习任务。书中对技术背景的解释清晰易懂, 同时也给出了详尽的 MATLAB 实例代码。

中文市场上已有不少针对 MATLAB 数值计算和仿真的专著, 但是专门针对在 MATLAB 中进行机器学习的专业书籍很少, 本书刚好在这个方面填充了市场的空白。我们在和很多客户的沟通过程中, 大家普遍反应对机器学习知其然, 却不知其所以然; 要把机器学习技术应用到具体工作中茫无头绪。这也是我们要翻译本书的主要动机, 希望能够通过这本关于 MATLAB 与机器学习的专著, 能够给有志于掌握机器学习原理的工程师和研究者提供一个扎实的台阶。

作者具有 20 多年 MATLAB 大型工业项目开发的实践经历和丰富的教学经验。书中全面涵盖了机器学习领域的关键技术内容, 原理阐释简洁清晰。应用实例则以独特的“问题-方法-步骤”的形式呈现给读者, 具有极强的针对性与实用性, 非常便于读者以问题驱动的方式快速有效地展开学习。另外, 本书从一个更加宽泛的角度 (自主系统) 入手, 在介绍算法本身之外, 也提供了对系统建模的原理性介绍, 帮助读者从机器学习在工程领域的具体应用中受到启发。

倍感荣幸有机会成为本书的译者。事实上, 翻译过程也是一次全新的学习过程, 重新梳理知识框架, 构建新的知识和技术的逻辑体系, 加深对技术演进的理解。所以, 我们向读者认真地推荐这本书, 辅以代码编写、调试与运行实践, 相信读者将会充分拓展、丰富自己的知识储备和技术领域, 利用强大的 MATLAB 工具进入机器学习技术的快速演进中。

ABOUT THE AUTHORS

作者简介

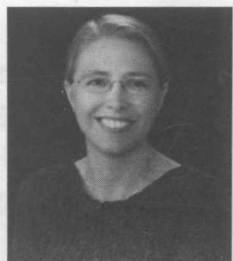
Michael Paluszek 是普林斯顿卫星系统 (PSS) 公司总裁, 该公司位于美国新泽西州普莱恩斯伯勒。Paluszek 先生于 1992 年创建了 PSS 公司, 主要业务是提供航空航天咨询服务。他使用 MATLAB 开发了地球同步通信卫星的控制系统和仿真系统 Indostar-1, 并于 1995 年推出了普林斯顿卫星系统公司的第一个商业 MATLAB 工具箱: 航天器控制工具箱。从那时起, 他已经先后为飞行器、潜水艇、机器人和核聚变推进系统等开发了工具箱和软件包, 并且形成了覆盖范围广泛的公司产品线。他目前正在领导一个美国陆军小型卫星精密姿态控制的研究合同项目, 并与普林斯顿等离子体物理实验室合作开发一个用于发电和太空推进的紧凑型核聚变反应堆。



在成立 PSS 公司之前, Paluszek 先生是位于新泽西州东温莎的通用电气公司 (GE) 宇航部门的工程师。在通用电气公司, 他设计了全球地球空间科学极地消旋平台控制系统, 并主导设计了 GPS IIR 姿态控制系统、Inmarsat-3 姿态控制系统和火星观测器 Delta-V 控制系统, 这些系统的控制设计都使用了 MATLAB。Paluszek 先生还致力于 DMSP 气象卫星姿态确定系统的研发。Paluszek 先生参与了超过 12 颗通信卫星的发射任务, 其中包括 GSTAR III 恢复任务, 第一次使用电推进器将卫星转移到作业轨道。在 Draper 实验室工作期间, Paluszek 先生负责航天飞机、空间站和海底导航等工作。他的空间站工作包括基于控制力矩陀螺仪系统的姿态控制设计。

Paluszek 先生获得了麻省理工学院的电气工程学士学位、航空航天学硕士和工程学位。他发表了很多论文, 拥有十多项美国专利。Paluszek 先生是 Apress 出版社出版的《MATLAB Recipes》一书的合著者。

Stephanie Thomas 是位于美国新泽西州普莱恩斯伯勒的普林斯顿卫星系统公司的副总裁。她分别于 1999 年和 2001 年从麻省理工学院获得航空航天学士学位和硕士学位。Thomas 女士于 1996 年在暑期实习期间加入 PSS 公司的 MATLAB 航天器控制工具箱开发项目, 自那以后就一直使用 MATLAB 进行航空航天分析。在近 20 年的 MATLAB 实践经历中, 她开发了许多软件工具, 包括用于航天器控制工具箱的太阳能帆板模块; 美国空军的近地轨道卫星操控工具



箱；用于 Prisma 卫星任务的碰撞监测 Simulink 模块；用 MATLAB 和 Java 编写的运载火箭分析工具。她开发了空间状态评估的新方法，例如用 MATLAB 和 C++ 两种语言实现的数值算法，用来评估任意两颗卫星之间的一般会合问题。Thomas 女士还为普林斯顿卫星系统公司的《Attitude and Orbit Control》教材编写做出了贡献，其中介绍了使用航天器控制工具箱 (SCT) 的案例，并编写了许多软件用户指南。她为来自澳大利亚、加拿大、巴西和泰国等不同国家的工程师进行了航天器控制工具箱培训，并为美国太空总署 (NASA)、美国空军和欧洲航天局等提供 MATLAB 咨询服务。Thomas 女士是 Apress 出版的《MATLAB Recipes》一书的合著者。2016 年，Thomas 女士因“核聚动力冥王星轨道探测器和登陆器”入选美国太空总署创新资助项目，被任命为美国太空总署 NIAC 研究员。

ABOUT THE REVIEWERS

技术审校者简介

Jonah Lissner 是一名研究员，在理论物理、电力工程、复杂系统、超材料、地球物理和计算理论等领域积极推进博士和理学博士计划、奖学金、应用项目和学术期刊出版物等。针对假设建构、理论学习、数学与公理建模以及解决抽象问题的测试，他在经验主义和科学理性方面具有强大的认知能力。他的博士论文、研究出版物和项目、简历、期刊、博客、小说和系统等内容都罗列在网站 <http://Lissnerresearch.weebly.com>。

Joseph Mueller 博士专攻控制系统和轨迹优化。在博士论文中，他为平流层飞艇开发了最佳上升轨迹。他的研究兴趣包括健壮性最优控制、自适应控制、应用优化和规划决策支持系统，以及实现机器人车辆自主运行的智能系统。

在 2014 年年初加入 SIFT 公司之前，**Mueller** 博士在普林斯顿卫星系统公司工作了 13 年。期间，他担任 NASA、美国空军、美国海军和美国导弹防御局（MDA）的 8 个小型企业创新研究合同项目的主要研究员。他开发了用于最佳引导和控制编队飞行航天器和高空飞艇的算法，并为美国国防部开发了一个通信卫星行动规划工具的培训课程。

2005 年，在参与 NASA Goddard 太空飞行中心的研究项目时，**Mueller** 博士开发了 MATLAB 编队飞行工具箱。作为一个商业产品，它现已用于 NASA、欧洲航天局，以及世界各地的大学和航空航天公司。

2006 年，**Mueller** 博士为瑞典 Prisma 卫星发射项目开发了安全轨道导航模式的算法和软件。自 2010 年发射以来，该项目已经成功地执行了两个航天器的编队飞行任务。

Mueller 博士还在明尼苏达大学（双子城校区）航空航天工程与力学系担任客座教授。

Derek Surka 在航空航天领域拥有超过 20 年的专业经验，专注于空间态势感知、引导、导航和控制，以及分布式系统自治和编队飞行。**Surka** 先生将他在天文动力学、数据融合、估计与控制系统，以及软件开发方面的专业知识应用于各种军用、民用和商业领域客户的 20 多颗卫星和有效载荷任务中。**Surka** 先生是一位积极的跑步爱好者和铁人三项运动员，还是美国前国家混合冰壶赛冠军。

前 言

机器学习正在众多学科中变得愈加重要，它应用于工程领域中的自动驾驶汽车技术和金融领域中的股市预测，而医疗专业人员则使用它来辅助诊断。虽然许多优秀的机器学习软件包可以通过商业购买和开源软件渠道获得，但深入理解其中隐藏的算法原理仍然是很有价值的。进而，自己动手编程来实现算法则会更加受益匪浅，因为这样不仅能够深入了解商业和开源软件包中的算法实现方法，还能掌握足够的背景知识来编写定制化的机器学习软件以实现特定的应用需求。

MATLAB 的起源正是基于这样的目的。最初，科学家们使用 FORTRAN 语言编写数值软件来进行矩阵运算。当时，用户必须通过“编写 - 编译 - 链接 - 执行”的过程来使用计算机程序，整个过程非常耗时，且极易出错。MATLAB 则为用户提供了一种脚本语言，用户只须编写很少的几行代码，立即执行，便可以解决许多问题。MATLAB 的内置可视化工具可以进一步帮助用户更好地理解计算结果。编写 MATLAB 程序比编写 FORTRAN 程序更为高效和充满乐趣。

本书旨在帮助用户利用 MATLAB 解决一系列宽泛的学习问题。本书包含两部分：第一部分包括第 1~3 章，介绍机器学习的背景知识，其中包括学习控制，其内容通常与机器智能并不紧密相关，在书中我们采用“自主学习”一词涵盖所有这些学科。本书第二部分包括第 4~12 章，展示了完整的 MATLAB 机器学习应用示例。第 4~6 章针对性地介绍了 MATLAB 的相关功能，使得机器学习算法非常易于实现。其余章节则给出了应用示例。每一章都提供了特定主题的技术背景和如何实现学习算法的思路。每个示例都由一系列 MATLAB 函数支持的 MATLAB 脚本来实现。

本书适用于信息领域中对机器学习感兴趣的技术人员和开发者，也适用于其他技术领域中对如何利用机器学习和 MATLAB 来解决专业领域问题感兴趣的技术人员。

CONTENTS

目 录

推荐序	2.3 学习控制	17
译者序	2.4 机器学习	18
作者简介	2.5 未来	19
技术审校者简介	参考文献	19
前言	第3章 机器学习软件	20
	3.1 自主学习软件	20
	3.2 商业化 MATLAB 软件	21
	3.2.1 MathWorks 公司产品	21
	3.2.2 普林斯顿卫星系统 产品	22
	3.3 MATLAB 开源资源	23
	3.3.1 深度学习工具箱	23
	3.3.2 深度神经网络	23
	3.3.3 MatConvNet	23
	3.4 机器学习工具	23
	3.4.1 R 语言	23
	3.4.2 scikit-learn	24
	3.4.3 LIBSVM	24
	3.5 优化工具	24
	3.5.1 LOQO	24
	3.5.2 SNOPT	25
	3.5.3 GLPK	25
	3.5.4 CVX	25
	3.5.5 SeDuMi	25
	3.5.6 YALMIP	26
	参考文献	26
第一部分 机器学习概论		
第1章 机器学习概述		2
1.1 引言		2
1.2 机器学习基础		3
1.2.1 数据		3
1.2.2 模型		3
1.2.3 训练		4
1.3 学习机		5
1.4 机器学习分类		6
1.5 自主学习方法		7
1.5.1 回归		8
1.5.2 神经网络		10
1.5.3 支持向量机		11
1.5.4 决策树		11
1.5.5 专家系统		12
参考文献		13
第2章 自主学习的历史		14
2.1 引言		14
2.2 人工智能		14

第二部分 机器学习的 MATLAB 实现

第 4 章 用于机器学习的

MATLAB 数据类型 28

4.1 MATLAB 数据类型概述 28

4.1.1 矩阵 28

4.1.2 元胞数组 29

4.1.3 数据结构 30

4.1.4 数值类型 31

4.1.5 图像 31

4.1.6 数据存储 33

4.1.7 Tall 数组 34

4.1.8 稀疏矩阵 35

4.1.9 表与分类数组 35

4.1.10 大型 MAT 文件 36

4.2 使用参数初始化数据结构 37

4.2.1 问题 37

4.2.2 方法 37

4.2.3 步骤 37

4.3 在图像数据存储上执行

mapreduce 39

4.3.1 问题 39

4.3.2 方法 39

4.3.3 步骤 39

总结 41

第 5 章 MATLAB 图形 42

5.1 二维线图 42

5.1.1 问题 42

5.1.2 方法 42

5.1.3 步骤 43

5.2 二维图形 47

5.2.1 问题 47

5.2.2 方法 47

5.2.3 步骤 47

5.3 定制二维图 51

5.3.1 问题 51

5.3.2 方法 51

5.3.3 步骤 51

5.4 三维盒子 57

5.4.1 问题 57

5.4.2 方法 57

5.4.3 步骤 57

5.5 用纹理绘制三维对象 59

5.5.1 问题 59

5.5.2 方法 59

5.5.3 步骤 59

5.6 三维图形 61

5.6.1 问题 61

5.6.2 方法 61

5.6.3 步骤 61

5.7 构建图形用户界面 62

5.7.1 问题 62

5.7.2 方法 62

5.7.3 步骤 63

总结 72

第 6 章 MATLAB 机器学习

示例 73

6.1 引言 73

6.2 机器学习 73

6.2.1 神经网络 73

6.2.2 面部识别 74

6.2.3 数据分类 74

6.3 控制 74

6.3.1 卡尔曼滤波器 74

6.3.2 自适应控制	75	7.8 测试神经网络	95
6.4 人工智能	76	7.8.1 问题	95
第7章 基于深度学习的		7.8.2 方法	95
面部识别	77	7.8.3 步骤	95
7.1 在线获取数据:用于训练神经		7.9 识别图像	97
网络	80	7.9.1 问题	97
7.1.1 问题	80	7.9.2 方法	97
7.1.2 方法	80	7.9.3 步骤	97
7.1.3 步骤	80	总结	98
7.2 生成神经网络的训练数据	80	参考文献	98
7.2.1 问题	80	第8章 数据分类	99
7.2.2 方法	80	8.1 生成分类测试数据	99
7.2.3 步骤	81	8.1.1 问题	99
7.3 卷积	85	8.1.2 方法	99
7.3.1 问题	85	8.1.3 步骤	99
7.3.2 方法	85	8.2 绘制决策树	102
7.3.3 步骤	85	8.2.1 问题	102
7.4 卷积层	87	8.2.2 方法	102
7.4.1 问题	87	8.2.3 步骤	102
7.4.2 方法	87	8.3 决策树的算法实现	106
7.4.3 步骤	87	8.3.1 问题	106
7.5 池化	89	8.3.2 方法	106
7.5.1 问题	89	8.3.3 步骤	106
7.5.2 方法	90	8.4 生成决策树	109
7.5.3 步骤	90	8.4.1 问题	109
7.6 全连接层	91	8.4.2 方法	109
7.6.1 问题	91	8.4.3 步骤	109
7.6.2 方法	91	8.5 手工创建决策树	114
7.6.3 步骤	91	8.5.1 问题	114
7.7 确定输出概率	93	8.5.2 方法	114
7.7.1 问题	93	8.5.3 步骤	114
7.7.2 方法	93	8.6 训练和测试决策树	118
7.7.3 步骤	93	8.6.1 问题	118

8.6.2	方法	118	10.1.3	步骤	156
8.6.3	步骤	119	10.1.4	传统卡尔曼滤波器	159
	总结	124	10.2	使用 UKF 进行状态估计	169
	参考文献	124	10.2.1	问题	169
第 9 章 基于神经网络的			10.2.2	方法	169
数字分类		125	10.2.3	步骤	169
9.1	生成带噪声的测试图像	125	10.3	使用 UKF 进行参数估计	176
9.1.1	问题	125	10.3.1	问题	176
9.1.2	方法	125	10.3.2	方法	176
9.1.3	步骤	126	10.3.3	步骤	176
9.2	创建神经网络工具箱	128	总结		183
9.2.1	问题	128	参考文献		183
9.2.2	方法	129	第 11 章 自适应控制		184
9.2.3	步骤	129	11.1	自调谐: 求振荡器频率	185
9.3	训练单一输出节点的神经网络	138	11.1.1	问题	186
9.3.1	问题	138	11.1.2	方法	186
9.3.2	方法	139	11.1.3	步骤	186
9.3.3	步骤	139	11.2	模型参考自适应控制	192
9.4	测试神经网络	143	11.2.1	创建方波输入	192
9.4.1	问题	143	11.2.2	实现模型参考自适应控制	194
9.4.2	方法	143	11.2.3	转子的 MRAC 系统实现	197
9.4.3	步骤	143	11.3	飞机的纵向控制	199
9.5	训练多输出节点的神经网络	144	11.3.1	编写飞机纵向运动的微分方程	200
9.5.1	问题	144	11.3.2	利用数值方法寻找平衡状态	206
9.5.2	方法	144	11.3.3	飞机的数值仿真	207
9.5.3	步骤	144	11.3.4	神经网络中对取值范围的限定和缩放	209
	总结	147	11.3.5	寻找学习控制的神经网络	210
	参考文献	148			
第 10 章 卡尔曼滤波器		149			
10.1	状态估计器	150			
10.1.1	问题	150			
10.1.2	方法	155			

11.3.6	枚举输入集合	213	12.1.3	方法	240
11.3.7	编写通用神经网络函数	216	12.2	汽车的自主传递控制	244
11.3.8	实现PID控制	219	12.2.1	问题	244
11.3.9	飞机俯仰角PID控制演示	224	12.2.2	方法	244
11.3.10	创建俯仰动力学的神经网络	228	12.2.3	步骤	244
11.3.11	非线性仿真中的控制器演示	230	12.3	汽车动力学	246
11.4	轮船驾驶:实现轮船驾驶控制的增益调度	232	12.3.1	问题	246
11.4.1	问题	232	12.3.2	步骤	246
11.4.2	方法	233	12.3.3	方法	248
11.4.3	步骤	233	12.4	汽车仿真与卡尔曼滤波器	250
总结		238	12.4.1	问题	250
参考文献		238	12.4.2	方法	250
第12章 自动驾驶		239	12.4.3	步骤	250
12.1	汽车雷达建模	239	12.5	雷达数据的MHT实现	257
12.1.1	问题	239	12.5.1	问题	257
12.1.2	步骤	239	12.5.2	方法	257
			12.5.3	步骤	260
			12.5.4	假设形成	268
			12.5.5	轨道剪枝	276
			总结		287
			参考文献		288

PART I

第一部分

机器学习概论

第1章 机器学习概述

第2章 自主学习的历史

第3章 机器学习软件

机器学习概述

1.1 引言

机器学习属于计算机科学的一个分支，它利用已有数据对未来数据做出预测或响应。它与模式识别、计算统计学和人工智能等领域密切相关。机器学习在诸如人脸识别、垃圾邮件过滤，以及那些不可行或甚至不可能通过编写算法来执行任务的领域中发挥着重要的作用。

例如，早期尝试垃圾邮件过滤时，由用户定义规则来确定什么是垃圾邮件。成功与否取决于用户是否能够正确识别将电子邮件归类为垃圾邮件的特定属性（例如发件人地址或主题关键字），以及用户在对规则进行细微调整上愿意投入的时间成本。但是因为垃圾邮件发送者在预测过滤规则方面并不会有什么困难，这种方法的效果有限。现代系统使用机器学习技术取得了更大的成功。我们大多数人现在已经熟悉将指定邮件标记为“垃圾邮件”或“非垃圾邮件”的概念，我们认为邮件系统可以快速学习这些电子邮件的哪些特征将其标识为垃圾邮件，并阻止这类邮件继续出现在收件箱中。这些特征可以是 IP 地址、邮件地址、邮件主题或正文中关键字的任意组合，以及各种匹配规则。请注意该示例中的机器学习如何以数据驱动方式在你接收电子邮件并标记它时自主地、不断地更新自身的学习规则。

那么在更泛化的意义上，机器学习是什么呢？机器学习可以意味着使用机器（计算机硬件与软件）从数据中获得知识，也意味着赋予机器从环境中学习的能力。数千年来机器已用于帮助人类。考虑一个简单的杠杆，它可以使使用岩石和一定长度的木头来构造，或者利用倾斜平面。这两台机器都能够完成有用的工作并且帮助人类，但它们并没有学习能力，因为它们都受自身的构建方式所限制。一旦建成，如果没有人类干预，它们就不能适应不断变化的需求。图 1-1 显示了早期不具备学习能力的简单机器。

这两台机器都能完成有用的工作，增强人类的能力。知识固化在它们的参数当中，也