

MATLAB Machine Learning Recipes
A Problem-Solution Approach
Second Edition

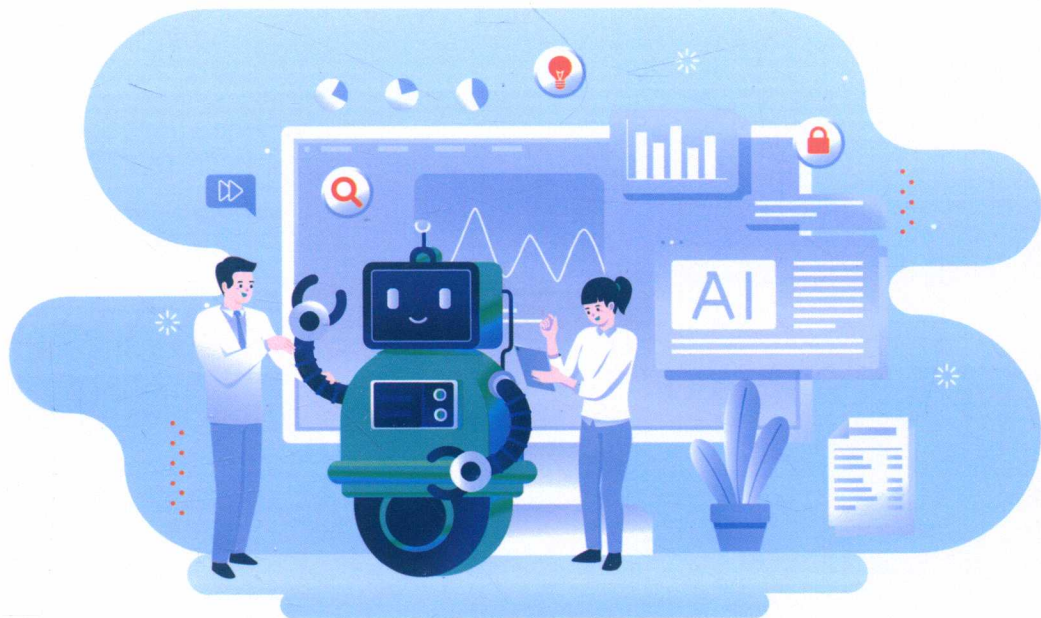
MATLAB机器学习

人工智能工程实践

(原书第2版)

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

陈建平 译



机械工业出版社
China Machine Press

智能系统与技术丛书

MATLAB Machine Learning Recipes
A Problem-Solution Approach
Second Edition


MATLAB机器学习

人工智能工程实践

(原书第2版)

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



 机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

MATLAB 机器学习: 人工智能工程实践 (原书第 2 版)/(美) 迈克尔·帕拉斯泽克 (Michael Paluszek), (美) 斯蒂芬妮·托马斯 (Stephanie Thomas) 著; 陈建平译. —北京: 机械工业出版社, 2020.2

(智能系统与技术丛书)

书名原文: MATLAB Machine Learning Recipes: A Problem-Solution Approach, Second Edition

ISBN 978-7-111-64677-8

I. M… II. ①迈… ②斯… ③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 (2020) 第 023834 号

本书版权登记号: 图字 01-2019-3878

First published in English under the title

MATLAB Machine Learning Recipes: A Problem-Solution Approach, Second Edition

by Michael Paluszek, Stephanie Thomas

Copyright © 2019 by Michael Paluszek and Stephanie Thomas

This edition has been translated and published under licence from

Apress Media, LLC, part of Springer Nature.

Chinese simplified language edition published by China Machine Press, Copyright © 2020.

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 机器学习: 人工智能工程实践 (原书第 2 版)

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: 冯秀泳

责任校对: 李秋荣

印 刷: 大厂回族自治县益利印刷有限公司

版 次: 2020 年 3 月第 1 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 19.75

书 号: ISBN 978-7-111-64677-8

定 价: 89.00 元

客服电话: (010) 88361066 88379833 68326294

投稿热线: (010) 88379604

华章网站: www.hzbook.com

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

内容简介

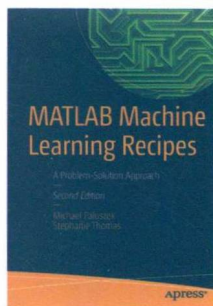
本书是关于在MATLAB中使用实例进行机器学习的综合指南。书中概述了人工智能与自动控制的历史；回顾了用于机器学习的商用软件包，并展示了它们如何应用于该领域；接着展示了如何使用MATLAB来解决机器学习问题，以及如何利用MATLAB图形技术来增强程序员对机器学习结果的理解。

本书随书提供了机器学习中若干重要问题的MATLAB完整解决方案，包括飞行控制、人脸识别、自动驾驶。书中所有的示例和应用程序都提供了完整的源代码。

机器学习包含大量的数学概念与理论解释。书中以清晰简洁的方式介绍了其中每个领域的数学知识，即使是不经常接触数学理论的读者也可以轻松理解。工程领域的读者会看到这些数学知识与他们已经了解的领域技术之间的密切联系，并将学习到新的技术。

译者简介

陈建平 MathWorks公司中国区的技术咨询专家，专注于工程大数据分析和高性能计算领域，拥有北京大学学士和硕士学位。加入MathWorks后，专注于通信系统和工程数据分析；深入探索工程数据在大数据领域的应用。他拥有十余年数值算法设计、实现以及大规模工程数据分析和建模的经验；尤其对MATLAB与不同编程语言，以及Hadoop和Spark等大数据架构的结合有较为深入的研究。



原书封面



华章图书

一本打开的书，
一扇开启的门，
通向科学殿堂的阶梯，
托起一流人才的基石。

Foreword 推荐序

随着机器学习和深度学习技术的迅猛发展，各种开发设计工具层出不穷。在众多可供选择的工具当中（包括像 Python、TensorFlow 这样的开源工具），MATLAB 在科学工程领域一直保持着独特的地位，是科学工作者和产品设计师开发机器学习应用的首选的可靠工具。主要原因在于以下三个方面：

- MATLAB 已经在科学和工程数据分析方面得到广泛的应用，是科研和工程工作者做数据分析的得力助手。
- MATLAB 提供的工具箱涵盖众多应用领域，包括计算金融、图像处理和计算机视觉、计算生物学、无线通信、汽车、航空航天和制造、自然语言处理，等等。
- MATLAB 提供 AI 开发的完整流程，包括数据产生、数据采集、数据预处理、特征工程、AI 算法开发、系统设计、系统部署，以及整合测试，可让行业专家快速掌握 AI 技术开发和产品设计的各个环节。

在 MathWorks 公司工作的 16 年中，我读过不少介绍 MATLAB 的图书，也浏览过很多关于机器学习方面的图书，但我感觉本书是为数不多的能真正把 MATLAB 的内在优势和机器学习的具体应用紧密结合，从解决问题的角度系统化地讲解相关知识的书籍之一。本书的内容完整地展现了 MATLAB 在机器学习中最重要和最有效的使用方式。作者以科研工作者所习惯的“问题—方法—步骤”的思维方式来解释每一个应用实例，既容易理解又有实用性。

本书译者陈建平拥有 20 多年的 MATLAB 的实战经验，对 MATLAB 的核心技术和应用技巧了如指掌。在 MathWorks 公司工作的 11 年中，他为很多应用领域的客户遇到的技术问题提供过很多解决方案。最近几年他致力于大数据和机器学习的应用，积累了丰富的实战经验和案例。他的翻译为这本书注入了新的价值。

不论你是行业专家，还是数据分析师，抑或是刚入门的机器学习爱好者，MATLAB 应该是你学习、研究和开发机器学习应用的必备工具，而本书应该是你用 MATLAB 开发机器学习的重要指南。

赵志宏

MathWorks 全球产品市场经理

译者序 *The Translator's Words*

从几年前的“人工智能距离我们有多远”，到现在的言必称深度学习，人工智能在过去的几年间获得了突飞猛进的发展。大量先进的生产力工具层出不穷，TensorFlow、PyTorch、MXNet 等各种深度学习框架让人目不暇接，人工智能看起来唾手可得。似乎只要掌握了框架的使用，我们就可以成为一个人工智能专家了。在概念上，机器学习替代了人工智能，深度学习替代了机器学习，深度学习好像成了人工智能的主流。在应用领域，大部分的成功案例都集中在机器视觉和自然语言处理上。当前流行的狭义的人工智能离实际工程渐行渐远。

实际上，人工智能距离工程应用就是一墙之隔。周志华在“西瓜书”中总结机器学习为“致力于研究如何通过计算的手段，利用经验来改善系统自身的性能”，从这个意义上来说，人工智能应该有更加广泛的意义和应用方向。我们希望能够从更加广泛的工程应用的角度来看待人工智能的应用和影响。人工智能首先应该是一种工程应用手段。

市场上已经有大量的书籍讨论如何掌握具体的机器学习框架，也有很多专门讨论机器学习和深度学习原理的书籍，它们大都探讨机器学习和深度学习本身，涉及如何将人工智能技术和工程应用相结合的书籍却如凤毛麟角。如何让人工智能成为现代工业的生产动力？我们之前和很多客户讨论机器学习和深度学习的时候，感兴趣者甚多，知道如何入手者甚少；知其然者多，不知其所以然者众。大家都认识到这是工业 4.0 的驱动力，却苦于不知如何跟传统工程问题结合到一起。这也是我想要翻译这本书的一个主要动机，希望对大家的工作有所帮助。

本书作者具有近 20 年 MATLAB 大型工程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，兼顾理论和实践。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。

本书的目的是利用 MATLAB 的强大功能来帮助用户从原理上学习机器学习，并用于解决具体工程问题，适用于每个对机器学习感兴趣的工程人员。

陈建平

2019 年 10 月

机器学习在每个工程学科中正变得越来越重要，比如：

1. 自动驾驶。机器学习几乎用于汽车控制系统的各个方面。
2. 等离子体物理学家借助机器学习的帮助来指导聚变反应堆的实验。实际上，TAE Systems 在指导聚变实验方面取得了巨大成功。普林斯顿等离子体物理实验室已将其用于国家球形环实验，以研究核聚变发电厂的可能的候选技术。
3. 在金融领域，它应用于预测股票市场。
4. 医疗专业人员将机器学习用于医疗诊断。
5. 执法部门将其用于面部识别，并在面部识别的辅助下解决了若干犯罪行为的定罪问题！
6. 美国太空总署的深空 1 号航天器使用了专家系统。
7. 用自适应控制系统操纵油轮。

还有许许多多其他的例子证明了机器学习的广泛应用。

虽然可以很方便地从商业来源和开源库获得许多优秀的机器学习软件包，但是去了解一下这些算法的工作原理依然很有价值。有机会编写自己的算法是非常值得的，它可以让你有机会深入体会和了解商业与开源软件包背后的原理。

MATLAB 的起源就是出于这个原因。当时，需要对矩阵进行操作的科学家使用 FORTRAN 来编写数值软件，要学会用计算机语言来完成编写—编译—链接—执行过程，耗时且容易出错。MATLAB 向用户提供了一种脚本语言，允许用户通过几行即时执行的脚本来解决许多问题。MATLAB 具有内置的可视化工具，可帮助用户更好地理解结果。显然编写 MATLAB 比编写 FORTRAN 更有效率和乐趣。

本书的目的是帮助所有用户利用 MATLAB 的强大功能来解决各种机器学习的问题，适用于每个对机器学习感兴趣的人。另外，本书还涵盖一些内容，可以让那些对其他技术领域感兴趣的人看到机器学习和 MATLAB 如何帮助他们解决专业领域的问题。

如何使用随书软件

随书软件附带一个 MATLAB 的工具箱，实现了本书涉及的所有例子。工具箱包含

1. MATLAB 函数
2. MATLAB 脚本
3. html 帮助

MATLAB 脚本实现了本书中的所有示例，而函数则封装了各种算法。许多函数都有内置的演示，只需在命令窗口中键入函数名称，它就会执行演示。该演示通常封装在子函数中。你可以为自己的演示复制此代码并将其粘贴到脚本中。例如，在命令窗口中键入函数名称 `PlotSet`，将出现图 1。

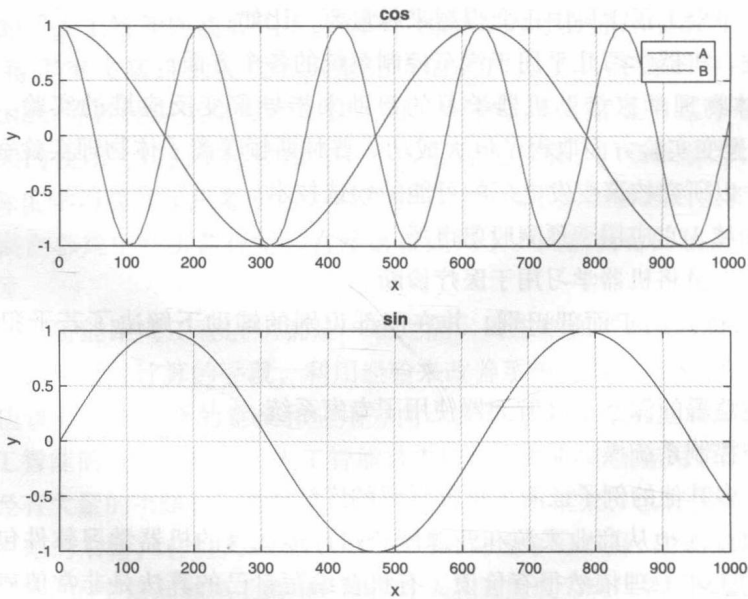


图 1 基于函数 `PlotSet.m` 的作图示例

打开函数源代码，你可以看到内置的演示子函数：

```

%%% PlotSet>Demo
function Demo

x = linspace(1,1000);
y = [sin(0.01*x);cos(0.01*x);cos(0.03*x)];
disp('PlotSet:_One_x_and_two_y_rows')
PlotSet(x, y, 'figure_title', 'PlotSet_Demo',...
        'plot_set',{[2 3], 1}, 'legend',{{'A' 'B'}}, {'plot_title',
        {'cos', 'sin'}});

```

你可以基于这些演示开始构建你自己的脚本。某些函数，比如用于数值积分的右侧函数（函数句柄作为参数），则不带演示子函数。如果键入函数名，会报错：

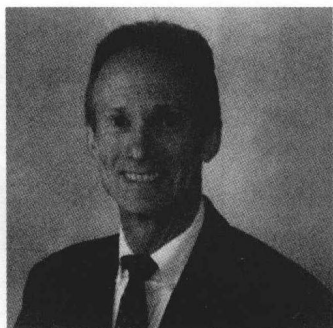
```
>> RHSAutomobileXY
Error using RHSAutomobileXY (line 17)
a built-in demo is not available.
```

该工具箱根据本书的章节进行组织，文件夹名称是 `Chapter_01`、`Chapter_02` 等。此外，还有一个 `general` 文件夹，其中包含支持工具箱其余部分的基础函数。你还需要开源包 GLPK (GNU Linear Programming Kit) 来运行一些代码。Nicolo Giorgetti 为 GLPK 编写了一个 MATLAB MEX 接口，该接口可在 SourceForge 上获得，并包含在此工具箱中。接口包括：

1. `glpk.m`
2. `glpkcc.mexmaci64` 或 `glpkcc.mexw64` 等
3. `GLPKTest.m`

该软件包可从 <https://sourceforge.net/projects/glpkmex/> 获取。第二项是为机器编译的 `glpkcc.cpp` 的 MEX 文件，分别对应 Mac 或 Windows 操作系统下的文件。请到 <https://www.gnu.org/software/glpk/> 获取 GLPK 库并将其安装在你的系统上。如果需要，也可以下载 GLPKMEX 源代码，并自行基于你的操作系统进行编译。

作者简介 *About the Authors*



Michael Paluszek 是普林斯顿卫星系统公司 (PSS) 总裁, 该公司位于美国新泽西州 Plainsboro。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 Indostar-1 地球同步通信卫星的控制系统和仿真系统, 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。

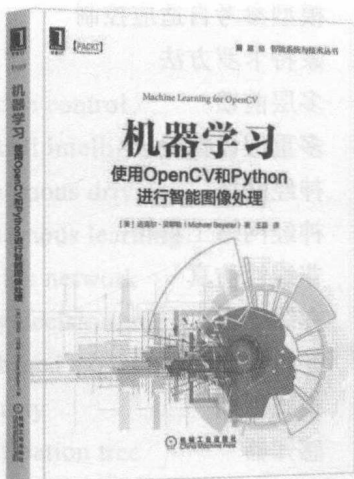
在成立普林斯顿卫星系统公司之前, Paluszek 先生是位于新泽西州 East Windsor 的通用电气公司 (GE) 宇航部门的工程师。在通用电气公司, 他设计了全球地球科学极地消旋平台控制系统, 并主导设计了 GPS IIR 姿态控制系统、Inmarsat-3 姿态控制系统和火星观测器 delta-V 控制系统, 这些系统的控制设计都使用了 MATLAB。Paluszek 先生还致力于 DMSP 气象卫星姿态确定系统的研发。Paluszek 先生参与了超过 12 颗通信卫星的发射任务, 其中包括 GSTAR III 恢复任务, 第一次使用电推进器将卫星转移到作业轨道。在 Draper 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 出版社出版的图书 *MATLAB Recipes* 和 *MATLAB Machine Learning* 的合著者。



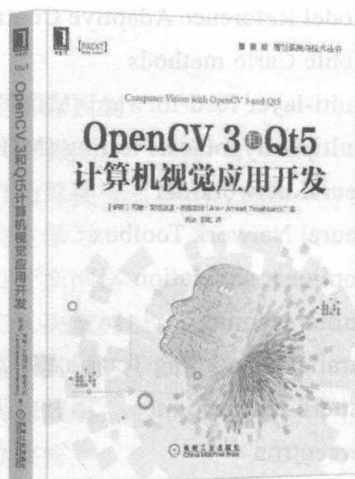
Stephanie Thomas 是位于美国新泽西州 Plainsboro 的普林斯顿卫星系统公司的副总裁。她于 1999 年和 2001 年分别从麻省理工学院获得航空航天学士学位和硕士学位。Thomas 女士于 1996 年在暑期实习期间加入普林斯顿卫星系统公司的 MATLAB 航天器控制工具箱开发项目，自那以后就一直使用 MATLAB 进行航空航天分析。在近 20 年的 MATLAB 实践经历中，她开发了许多软件工具：用于航天器控制工具箱的太阳能帆板模块，美国空军的近地轨道卫星操控工具箱，用于 Prisma 卫星任务的碰撞监测 Simulink 模块，用 MATLAB 和 Java 编写的运载火箭分析工具。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《姿态和轨道控制》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、美国空军和欧洲航天局等提供 MATLAB 咨询服务。Thomas 女士是 Apress 出版的图书 *MATLAB Recipes* 和 *MATLAB Machine Learning* 的合著者。2016 年，Thomas 女士因“核聚动力冥王星轨道探测和登陆器”项目，被任命为美国太空总署 NIAC 研究员。

推荐阅读



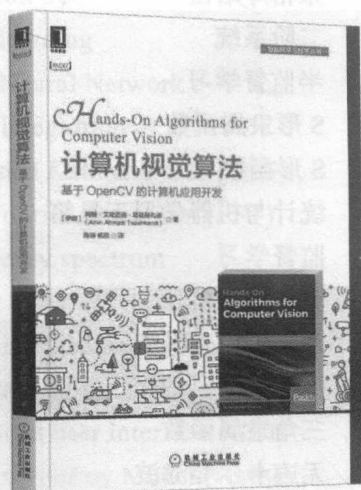
机器学习: 使用OpenCV和Python进行智能图像处理

作者: Michael Beyeler ISBN: 978-7-111-61151-6 定价: 69.00元



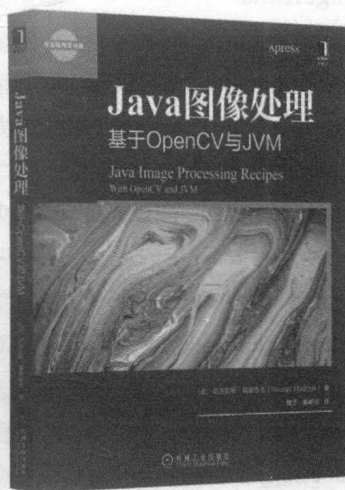
OpenCV 3和Qt5计算机视觉应用开发

作者: Amin Ahmaditazehkandi ISBN: 978-7-111-61470-8 定价: 89.00元



计算机视觉算法: 基于OpenCV的计算机应用开发

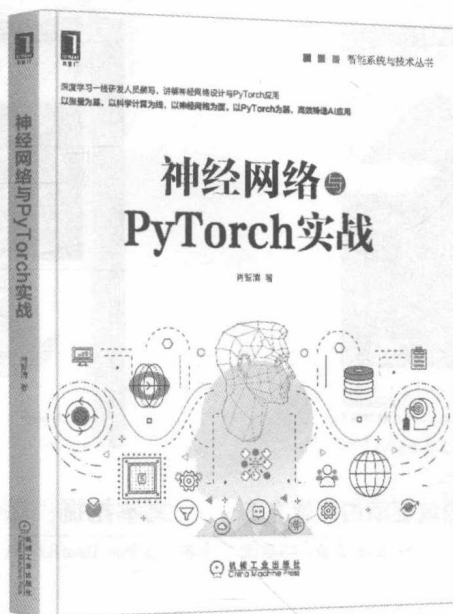
作者: Amin Ahmadi 等 ISBN: 978-7-111-62315-1 定价: 69.00元



Java图像处理: 基于OpenCV与JVM

作者: Nicolas Modrzyk ISBN: 978-7-111-62388-5 定价: 99.00元

推荐阅读

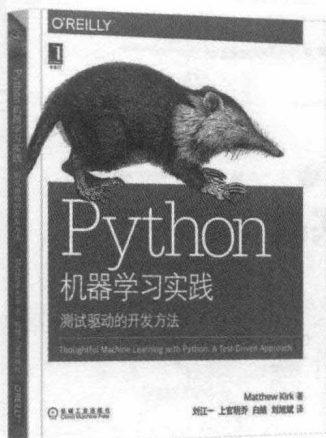


神经网络与PyTorch实战

书号：978-7-111-60577-5 作者：肖智清 定价：59.00元

**深度学习一线研发人员撰写，讲解神经网络设计与PyTorch应用
以张量为基，以科学计算为线，以神经网络为面，以PyTorch为器，高效铸造AI应用**

推荐阅读



Python机器学习实践：测试驱动的开发方法

作者：Matthew Kirk ISBN：978-7-111-58166-6 定价：59.00元



文本挖掘：基于R语言的整洁工具

作者：Julia Silge, David Robinson ISBN：978-7-111-58855-9 定价：59.00元



TensorFlow学习指南：深度学习系统构建详解

作者：Tom Hope, Yehezkel S. Resheff, Itay Lieder ISBN：978-7-111-60072-5 定价：69.00元



算法技术手册（原书第2版）

作者：George T. Heineman等 ISBN：978-7-111-56222-1 定价：89.00元

目 录 Contents

推荐序	第3章 MATLAB 作图	39
译者序	3.1 二维线图	39
前言	3.2 通用二维作图	42
作者简介	3.3 定制二维图表	44
	3.4 三维盒子	45
	3.5 用纹理绘制三维对象	47
	3.6 通用三维作图	49
	3.7 构建图形用户界面	51
	3.8 柱状图动画	56
	3.9 画一个机器人	59
	3.10 小结	61
第1章 机器学习概述	第4章 卡尔曼滤波	63
1.1 引言	4.1 用线性卡尔曼滤波器实现的状态估计器	64
1.2 机器学习基础	4.2 使用扩展卡尔曼滤波器进行状态估计	80
1.3 学习机	4.3 使用无迹卡尔曼滤波器进行状态估计	84
1.4 机器学习分类体系	4.4 使用无迹卡尔曼滤波器进行参数估计	90
1.5 控制	4.5 小结	93
1.6 自主学习方法	第5章 自适应控制	95
1.7 人工智能	5.1 自调谐: 振荡器建模	96
1.8 小结	5.2 自调谐: 调校振荡器	97
	5.3 模型参考自适应控制的实现	102
第2章 用于机器学习的 MATLAB 数据类型		
2.1 MATLAB 数据类型概述		
2.2 使用参数初始化数据结构		
2.3 在图像 datastore 上执行 mapreduce		
2.4 从文件中创建表格		
2.5 处理表格数据		
2.6 使用 MATLAB 字符串		
2.7 小结		