



# 机器学习入门到实战

## MATLAB 实践应用

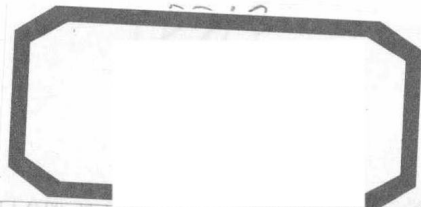
◎ 冷雨泉 张会文 张伟 等 著

### 本书知识体系

- k近邻算法
- 线性回归
- k均值算法
- 高斯混合模型
- 决策树
- 逻辑回归
- 期望最大化算法
- DBSCAN算法
- 支持向量机
- 神经网络
- k中心点算法
- 策略迭代和值迭代
- 朴素贝叶斯分类
- AdaBoost算法
- Apriori算法
- SARSA算法和Q学习算法



清华大学出版社



# 机器学习入门到实战

## MATLAB 实践应用

◎ 冷雨泉 张会文 张伟 等 著

清华大学出版社  
北京

## 内 容 简 介

本书主要介绍经典的机器学习算法的原理及改进,以及 MATLAB 的实例实现。本书内容分为三部分。第一部分(第 1 章)是机器学习概念篇,介绍机器学习的相关概念,并且对机器学习的各类算法进行分类,以便读者对机器学习的知识框架有一个整体的了解,从而在后续的学习中更容易接受机器学习涉及的各类算法。第二部分(第 2 章、第 3 章)是 MATLAB 机器学习基础篇,介绍 MATLAB 的基本使用方法,以及 MATLAB 集成的机器学习工具箱。MATLAB 易上手的特点让使用者将更多的精力专注于算法开发与使用,而不是搭建算法实现开发平台。第三部分(第 4 章~第 19 章)是机器学习算法与 MATLAB 实践篇,对监督学习、无/非监督学习、强化学习三大类常用算法进行逐个讲解,包括机器学习算法原理、算法优缺点、算法的实例解释以及 MATLAB 的实践应用。

本书适合以下读者:对人工智能、机器学习感兴趣的读者;希望用机器学习完成设计的计算机或电子信息专业学生;准备开设机器学习、深度学习实践课的授课老师;学习过 C 语言,且希望进一步提升编程水平的开发者;刚从事机器学习、语音、机器视觉、智能机器人研发的算法工程师。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

## 图书在版编目(CIP)数据

机器学习入门到实战:MATLAB 实践应用/冷雨泉等著. —北京:清华大学出版社,2019  
(2019.5重印)

(大数据与人工智能技术丛书)

ISBN 978-7-302-49514-7

I. ①机… II. ①冷… III. ①Matlab 软件—应用—机器学习 IV. ①TP181

中国版本图书馆 CIP 数据核字(2018)第 029354 号

策划编辑:魏江江  
责任编辑:王冰飞  
封面设计:刘 键  
责任校对:时翠兰  
责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市龙大印装有限公司

经 销:全国新华书店

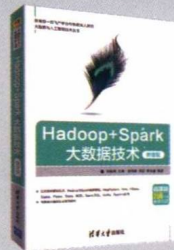
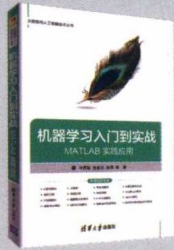
开 本:185mm×260mm 印 张:17.5 插 页:4 字 数:370 千字

版 次:2019 年 3 月第 1 版 印 次:2019 年 5 月第 2 次印刷

印 数:2001~3500

定 价:59.00 元

产品编号:075373-01



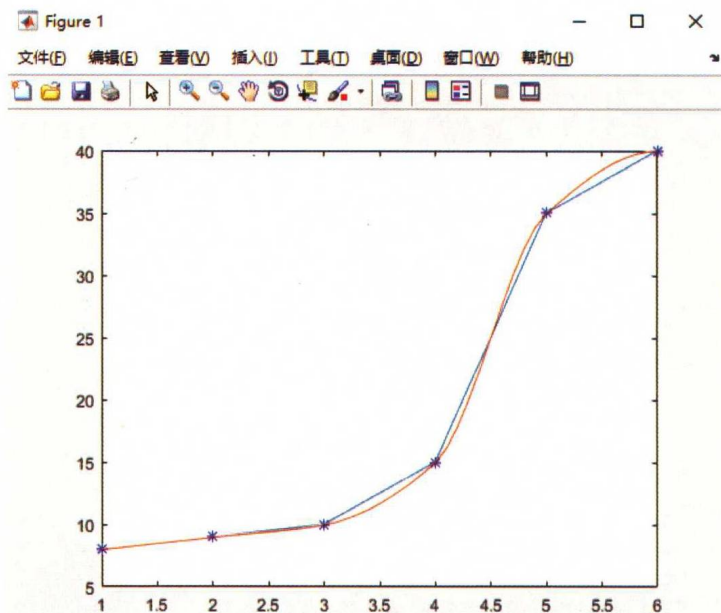


图 2.4 利用 Plot 函数绘图

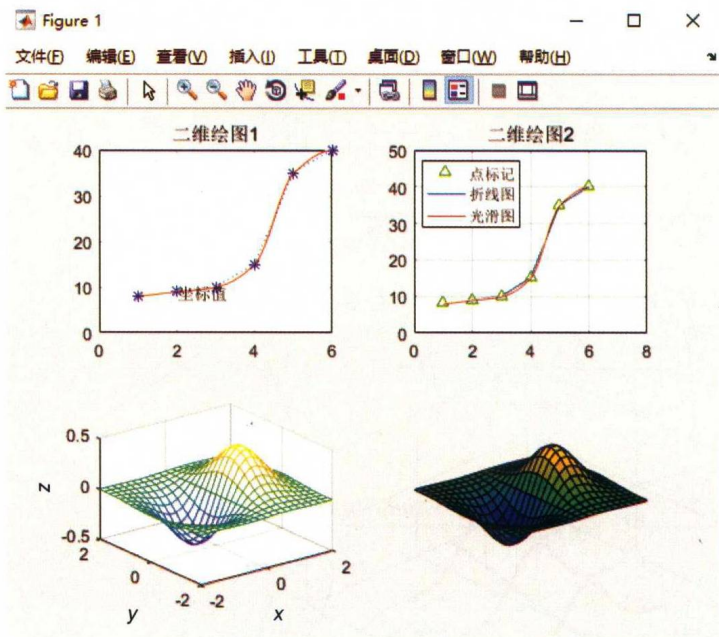


图 2.7 综合绘图实例

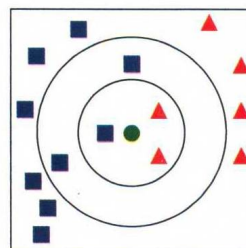


图 4.1  $k$  近邻算法实例图

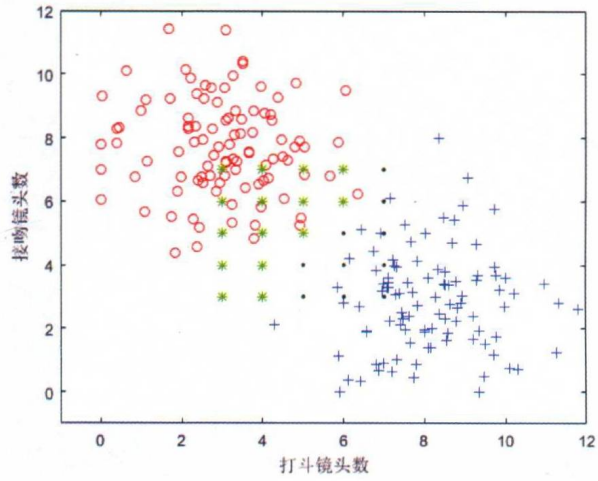


图 4.5 KNN 算法电影类型分类

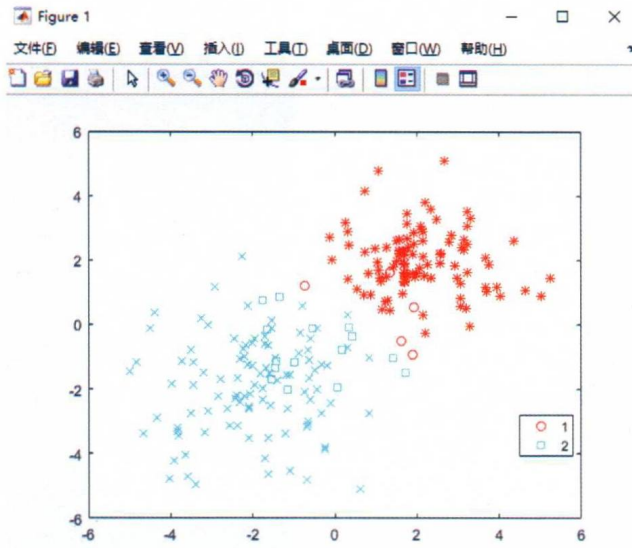


图 4.6 knnclassify 函数的 KNN 分类

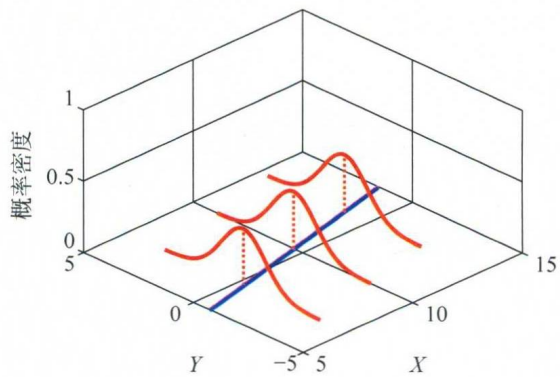


图 9.4 线性回归模型的实例结果

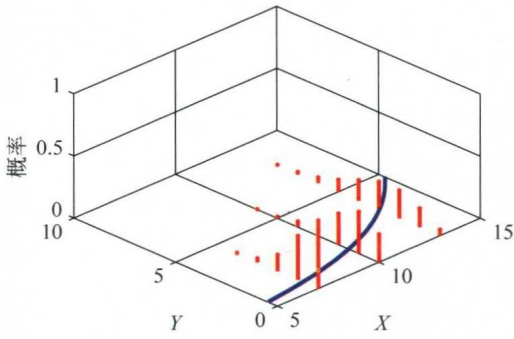


图 9.5 广义线性模型拟合指数曲线的实例

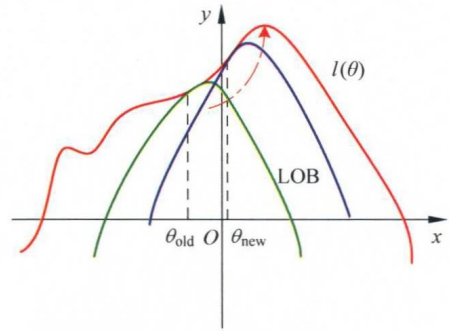


图 13.1 EM 算法的思想

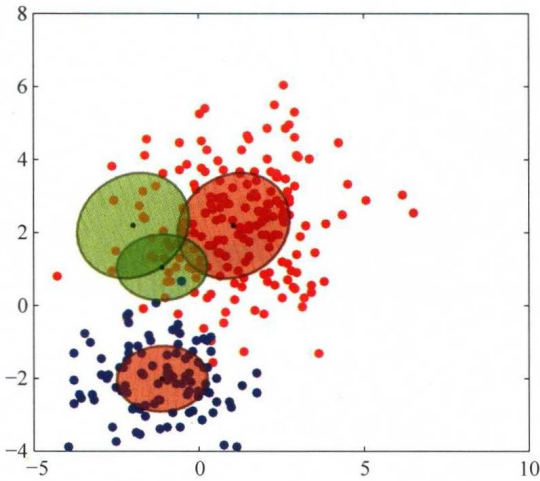


图 13.3 EM 算法用于 GMM 的示例

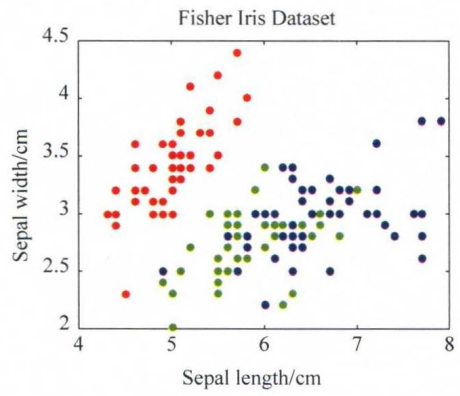


图 16.2 鸢尾花数据集

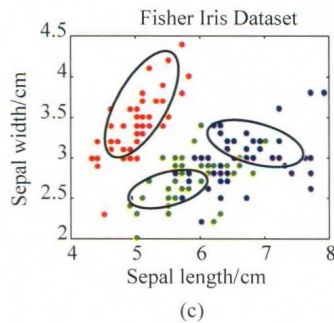
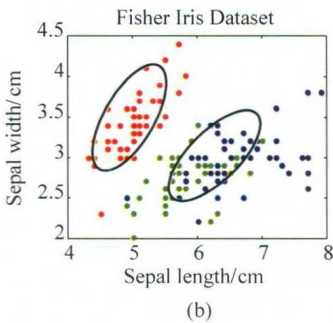
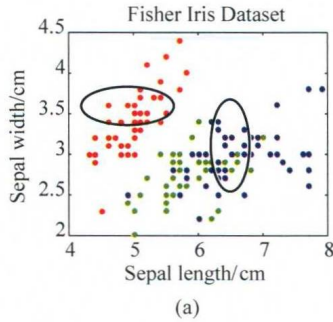


图 16.3 不同的聚类方式(黑色虚线椭圆代表一个类)

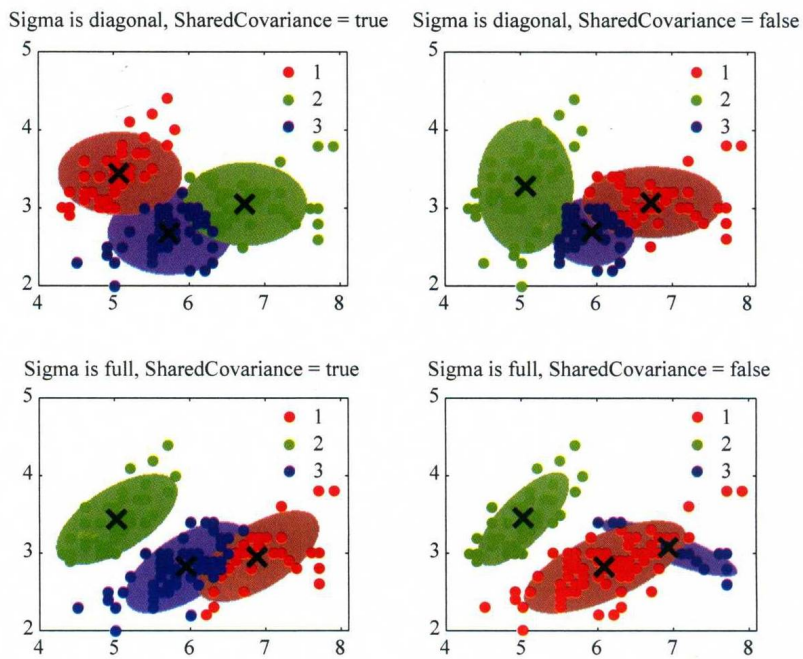


图 16.4 GMM 在 Fisher Iris 数据集上的聚类效果

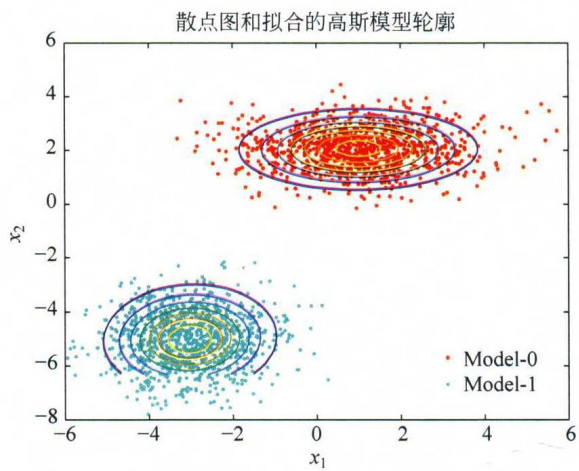


图 16.6 拟合高斯混合模型实例

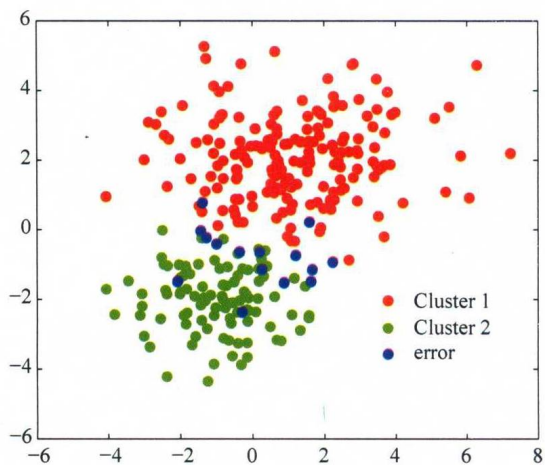


图 16.9 利用 GMM 聚类

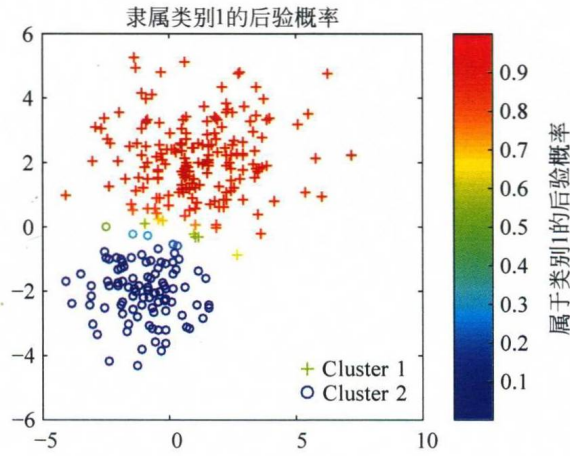


图 16.10 类别 1 的隶属度

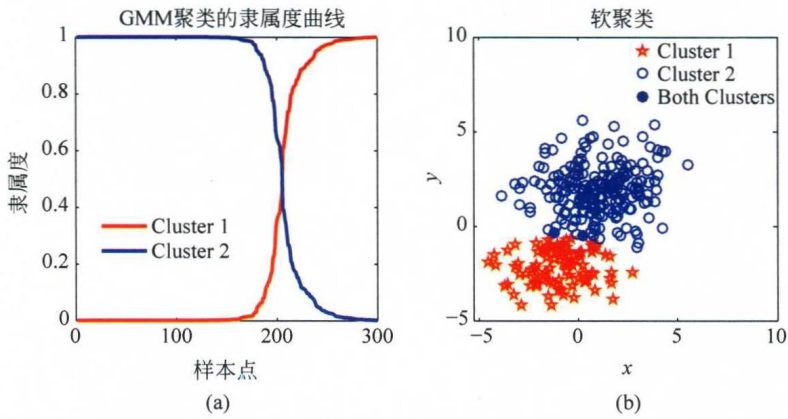


图 16.12 隶属度值和软聚类

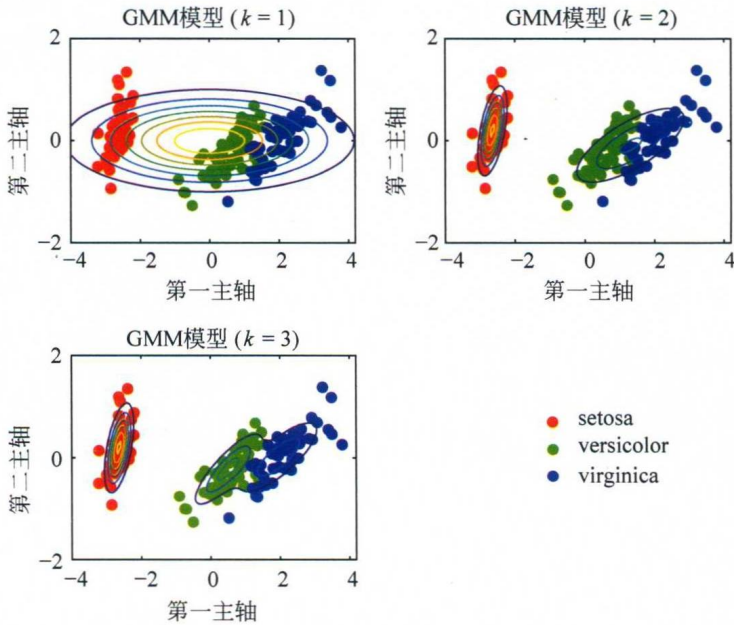


图 16.14 GMM 中使用不同  $k$  的效果

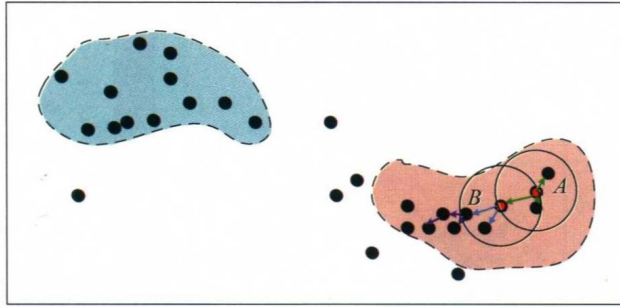


图 17.3 DBSCAN 算法聚类的过程

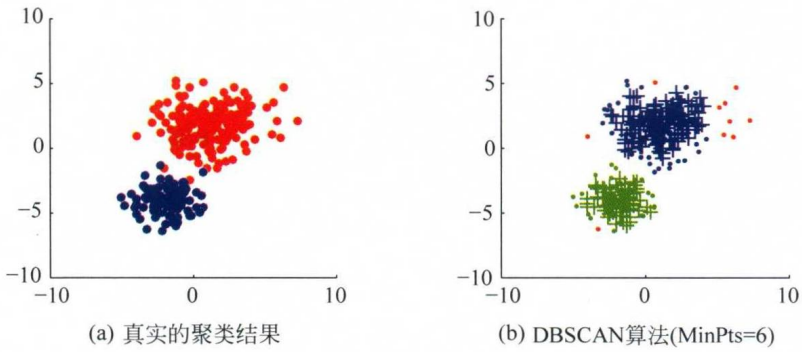


图 17.4 DBSCAN(MinPts=6)预测与真实结果对比

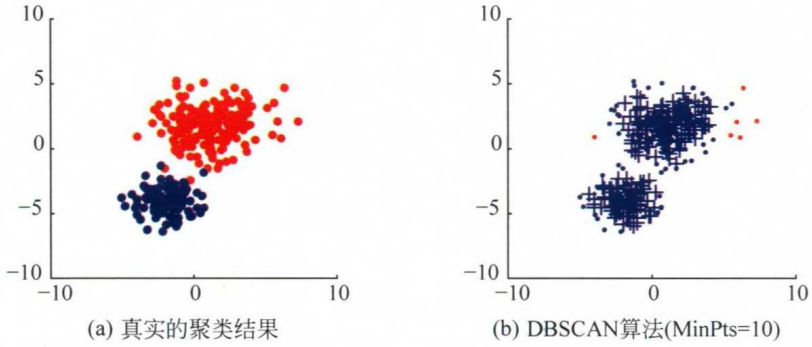


图 17.5 DBSCAN(MinPts=10)预测与真实结果对比

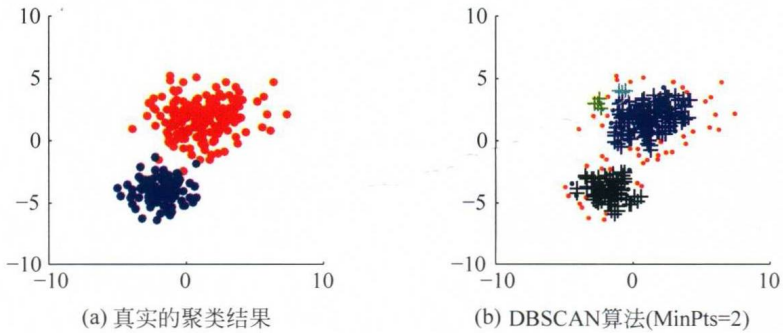


图 17.6 DBSCAN(MinPts=2)预测与真实结果对比

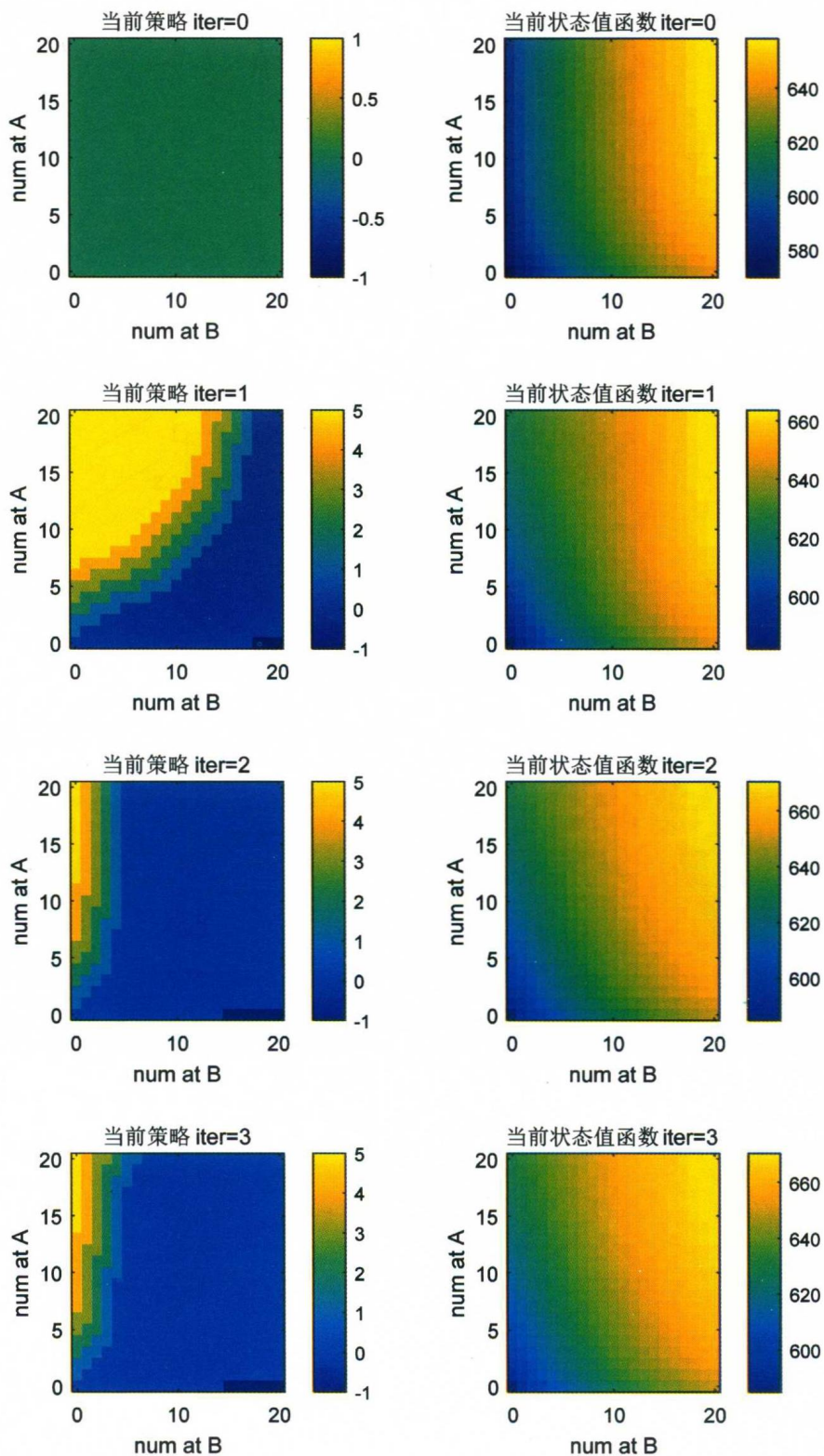


图 18.6 每次迭代时的策略和对应的值函数

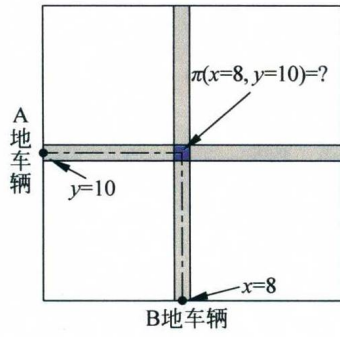


图 18.7 最优策略图的解释

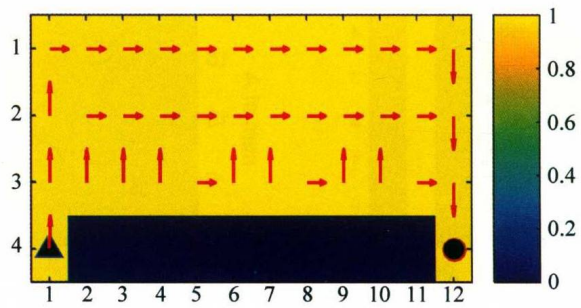


图 19.3 SARSA 算法得到的策略

# 前言

## FOREWORD

近年来,随着计算机技术及互联网技术的发展,人工智能技术也取得了重要的突破。作为人工智能的核心技术,机器学习已经广泛应用于各行各业中,如图像识别、语言识别、文本分类、智能推荐、网络安全等。未来,伴随着信息技术的进一步发展,机器学习技术将会更加深入地应用到生产、生活的方方面面。

目前,机器学习技术正处于朝阳时期,对于从事机器学习的研究人员来说,应感到荣幸和骄傲,因为能够在对的时间从事最热门的技术研究。对于有志于或有兴趣从事机器学习的研究人员而言,首先应知道,现阶段对这方面人才的需求远远大于供给,同时,这一技术会伴随着信息化技术一直发展下去。其次,在学习之初,不要被大量的数学公式吓得退避三舍,而应明白,在大多数情况下,尤其是应用层面,机器学习仅是一种实现技术要求的工具,需要了解各类算法的优势、劣势及有效使用的方法,无须详尽地了解各种机器学习算法的细枝末节。正如计算机内部运行机制极其复杂,大多数人每天都会使用,但却无须了解 CPU 和内存存在每一时刻的具体运行过程。

本书是作者在多年机器学习及工作经验的基础上,对大量的网络资源、论文和相关书籍进行总结、整理、分析后编写的。全书共分为三部分,分别为机器学习概念篇、MATLAB 机器学习基础篇、机器学习算法与 MATLAB 实践篇。

本书各章内容简介如下。

第 1 章主要介绍机器学习中的基本概念、术语等,机器学习算法分类框架,机器学习算法实现的基本流程,以及机器学习中数据预处理的基本方法。

第 2 章主要介绍 MATLAB 软件的基本操作和使用方法,包括矩阵运算、m 文件编写、流程控制语句编写、绘图及文件的导入与导出。

第 3 章主要介绍 MATLAB 机器学习工具箱中的分类学习器应用程序 (Classification Learner App) 的使用方法,通过安德森鸢尾花卉数据集的实例,具体介绍使用方法和流程。

第 4 章介绍分类回归算法中的  $k$  近邻算法 (KNN) 的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 5 章介绍分类回归算法中的决策树 (Decision Tree) 的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 6 章介绍分类回归算法中的支持向量机 (SVM) 的算法原理、算法实现步骤、算法

特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 7 章介绍分类回归算法中的朴素贝叶斯(Naive Bayes, NB)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 8 章介绍分类回归算法中的线性回归(Line Regression)的算法原理、算法实现步骤、多元线性回归原理,以及通过 MATLAB 进行实例的算法编写与详解。

第 9 章介绍分类回归算法中的逻辑回归(Logistic Regression)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 10 章介绍分类回归算法中的神经网络(Artificial Neural Networks, ANN)的算法原理、算法实现步骤、算法特点、算法拓展,以及通过 MATLAB 进行实例的算法编写与详解。

第 11 章介绍分类回归算法中的 AdaBoost 算法的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 12 章介绍聚类算法中的  $k$  均值算法( $k$ -means)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 13 章介绍聚类算法中的期望最大化算法(EM)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 14 章介绍聚类算法中的  $k$  中心点算法( $k$ -medoids)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 15 章介绍聚类算法中的关联规则挖掘的 Apriori 算法的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 16 章介绍聚类算法中的高斯混合模型(GMM)的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 17 章介绍聚类算法中的 DBSCAN 算法的算法原理、算法实现步骤、算法特点、算法改进,以及通过 MATLAB 进行实例的算法编写与详解。

第 18 章介绍强化学习算法中的策略迭代和值迭代的算法原理、算法实现步骤,以及通过 MATLAB 进行实例的算法编写与详解。

第 19 章介绍强化学习算法中的 SARSA 算法和 Q 学习算法的算法原理、算法实现步骤,以及通过 MATLAB 进行实例的算法编写与详解。

本书的出版得到了清华大学出版社图书出版基金的资助和出版社工作人员的大力支持,作者在此表示衷心的感谢。此外,学术界、产业界同仁们的不断探索,才推动机器学习技术走到今天,本书的完成得力于此,编者在此一并表示感谢。本书由冷雨泉、张会文、张伟著,其他参与编写的作者还有付明亮、韩小宁、秦晓成、张会彬,排名不分先后。

本书适合以下读者:对人工智能、机器学习感兴趣的读者;希望用机器学习完成设计的计算机或电子信息专业学生;准备开设机器学习、深度学习实践课的授课老师;学习过 C 语言,且希望进一步提升编程水平的开发者;刚从事机器学习、语音、机器视觉、智能机器人研发的算法工程师。

一方面,机器学习内容极为庞大和复杂,存在大量的交叉算法,且依据应用领域的不同,不同的算法也会有不同的表现;另一方面,机器学习领域发展极其迅速,不断取得新的

研究成果。因此,作者只能尽力将现有机器学习的框架关系及主要算法原理及其实现展现给读者,以起到抛砖引玉的作用,给予机器学习的初学者一定的指导。读者在后期的机器学习中,需要阅读大量的文献,并在实践中进行摸索。

由于作者学识有限,疏漏和不当之处在所难免,敬请读者和同行们给予批评指正(ML\_matlab@163.com)。读者如有兴趣,可加入机器学习互动QQ群446360728,进行交流,共同进步。

作者

2018年10月

## 第一部分 机器学习概念篇

●第1章 机器学习基础	3
1.1 机器学习概述	3
1.1.1 机器学习的概念	3
1.1.2 机器学习的发展史	4
1.1.3 机器学习的用途	6
1.1.4 机器学习、数据挖掘及人工智能的关系	9
1.2 机器学习基本术语	9
1.3 机器学习任务及算法分类	8
1.4 如何学习和运用机器学习	9
1.4.1 软件平台的选择	9
1.4.2 机器学习应用实现流程	11
1.5 数据预处理	12
1.5.1 数据初步选取	13
1.5.2 数据清理	13
1.5.3 数据集成	14
1.5.4 数据变换	11
1.5.5 数据归约	15
参考文献	12

## 第二部分 MATLAB 机器学习基础篇

●第2章 MATLAB 基础入门	19
2.1 MATLAB 界面介绍	19
2.2 矩阵赋值与运算	20
2.3 m 文件及函数实现与调用	23

## 目 录

4.2.1	4.2.1 决策树	4.2
4.2.2	4.2.2 支持向量机	4.2
4.2.3	4.2.3 分类回归树	4.2
CONTENTS	CONTENTS	
5.3	5.3 决策树算法MATLAB实践	5.3
参考文献	参考文献	
第6章	支持向量机	77
6.1	6.1 支持向量机算法原理	77
6.1.1	6.1.1 支持向量机	77
6.1.2	6.1.2 支持向量机算法及推广	78
第1章	机器学习基础	3
1.1	1.1 机器学习概述	3
1.1.1	1.1.1 机器学习的概念	3
1.1.2	1.1.2 机器学习的发展史	4
1.1.3	1.1.3 机器学习的用途	5
1.1.4	1.1.4 机器学习、数据挖掘及人工智能的关系	5
1.2	1.2 机器学习基本术语	6
1.3	1.3 机器学习任务及算法分类	8
1.4	1.4 如何学习和运用机器学习	9
1.4.1	1.4.1 软件平台的选择	9
1.4.2	1.4.2 机器学习应用实现流程	11
1.5	1.5 数据预处理	12
1.5.1	1.5.1 数据初步选取	13
1.5.2	1.5.2 数据清理	13
1.5.3	1.5.3 数据集成	13
1.5.4	1.5.4 数据变换	14
1.5.5	1.5.5 数据归约	15
参考文献	参考文献	15
第2章	MATLAB 机器学习基础篇	
第2章	MATLAB 基础入门	19
2.1	2.1 MATLAB 界面介绍	19
2.2	2.2 矩阵赋值与运算	20
2.3	2.3 m 文件及函数实现与调用	23

2.4	基本流程控制语句	24
2.5	基本绘图方法	26
2.5.1	二维绘图函数的基本用法	26
2.5.2	三维绘图函数的基本用法	28
2.5.3	颜色与形状参数列表	29
2.5.4	图形窗口分割与坐标轴	30
2.6	数据文件导入与导出	32
	参考文献	33
<b>●第3章 MATLAB 机器学习工具箱</b> 34		
3.1	工具箱简介	34
3.2	分类学习器基本操作流程	35
3.3	分类学习器算法优化与选择	40
3.3.1	特征选择	40
3.3.2	选择分类器算法	41
3.4	工具箱分类学习实例	42
	参考文献	49
<b>第三部分 机器学习算法与 MATLAB 实践篇</b>		
<b>●第4章 k 近邻算法</b> 53		
4.1	k 近邻算法原理	53
4.1.1	k 近邻算法实例解释	53
4.1.2	k 近邻算法的特点	54
4.2	基于 k 近邻算法的算法改进	55
4.2.1	快速 KNN 算法	56
4.2.2	k-d 树 KNN 算法	57
4.3	k 近邻算法的 MATLAB 实践	58
	参考文献	63
<b>●第5章 决策树</b> 64		
5.1	决策树算法原理	64
5.1.1	决策树算法基本原理	64
5.1.2	决策树算法的特点	66
5.1.3	决策树剪枝	66
5.1.4	分类决策树与回归决策树	68
5.2	基于决策树算法的算法改进	69