Broadview®
www.broadview.com.cn

MACHINE LEARNING ALGORITHM

深入分析机器学习中的常用算法，兼顾算法、理论与实践，帮助读者快速掌握算法精髓！

Python 机器学习算法

赵志勇◎著

探索数据的内在价值，
洞悉人工智能背后的技术！

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

Python 机器学习算法

安阳师范学院图书馆 藏书

赵志勇◎著

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

内 容 简 介

本书是一本机器学习入门读物，注重理论与实践的结合。全书主要包括 6 个部分，每个部分均以典型的机器学习算法为例，从算法原理出发，由浅入深，详细介绍算法的理论，并配合目前流行的 Python 语言，从零开始，实现每一个算法，以加强对机器学习算法理论的理解、增强实际的算法实践能力，最终达到熟练掌握每一个算法的目的。与其他机器学习类图书相比，本书同时包含算法理论的介绍和算法的实践，以理论支撑实践，同时，又将复杂、枯燥的理论用简单易懂的形式表达出来，促进对理论的理解。

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

Python 机器学习算法 / 赵志勇著. —北京：电子工业出版社，2017.7
ISBN 978-7-121-31319-6

Ⅰ.①P… Ⅱ.①赵… Ⅲ.①软件工具－程序设计 ②机器学习 Ⅳ.①TP311.561 ②TP181

中国版本图书馆 CIP 数据核字(2017)第 072742 号

策划编辑：符隆美
责任编辑：徐津平
印　　刷：北京季蜂印刷有限公司
装　　订：北京季蜂印刷有限公司
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：720×1000　1/16　印张：22.75　字数：426 千字
版　　次：2017 年 7 月第 1 版
印　　次：2017 年 7 月第 1 次印刷
定　　价：69.00 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。

推荐序

志勇是我在新浪微博的同事，刚来的时候坐我的旁边。记得当时志勇喜欢把看过的论文的重点部分剪下来粘到自己的笔记本上，并用五颜六色的笔标注，后来还知道志勇平时会写博客来记录自己在算法学习和实践中的心得。由此可见，志勇是一个非常认真且善于归纳总结的人。2016 年志勇告诉我他在写这样一本书的时候，我深以为然，感觉正确的人做了一件正确的事。

志勇的书快完成的时候邀请我写序，这对我绝对是个挑战。幸运的是，书中的内容是我所熟悉的，仿佛是发生在自己身边的事。写作风格也和志勇平时交流时一致。所以，我可以从故事参与者的角度去介绍一下这本书。

说到机器学习算法，这两年可谓蓬勃发展。AlphaGo 战胜世界围棋冠军李世石已经成了大家茶余饭后的谈资，无人驾驶汽车是资本竞相追逐的万亿级市场，这些都源于数据收集能力、计算能力的提升，以及智能设备的普及。机器学习也已经在我们身边一些领域取得了成功，例如，现在已经获得上亿用户使用的今日头条。今日头条通过抓取众多媒体的资讯，利用机器学习算法推荐给用户，从而做到了资讯量大、更新快、更加个性化。

作为一个互联网从业者，更是感觉到机器学习算法已经融入到越来越多的产品和功能中。我曾经在一次交流中得知，某个应用为减少用户输入，用了一个团队的力量做输入选项的推荐。这在以前是不曾出现过的，以前的网站更多的是增加功能，让用户选择。现在更多的是推荐给用户，帮助用户选择。这里面既有移动应用屏幕小、操作复杂的原因，也有互联网公司越来越重视用户体验的原因。所以，在此要恭喜这本书的读者，你们选择了一个前途光明的行业。

机器学习算法比较典型的应用是推荐、广告和搜索。我们利用协同过滤技术来推荐商品、利用逻辑回归技术来做点击率的预测、利用分类技术来识别“垃圾”网页等。学好、用好每一种算法都很困难，需要掌握背后的理论基础，以及进行大量的实践，否则就会浮于表面，模仿他人，不能根据自己的业务做出合理的选择。

而市场上的书通常要么只是一些概要性质的介绍，要么是偏向实战，理论基础介绍得比较少。本书是少有的两者兼具的书，每一种算法都先介绍数学基础，再用 Python 代码做简单版本的实现，并且算法之间循序渐进，层层深入，读来如沐春风。

本书介绍了 LR、FM、SVM、协同过滤、矩阵分解等推荐和广告领域常用算法，有很强的实用性。深度学习更是近期主流互联网公司研究的热门领域。无论对机器学习的初学者还是已经具备一些项目经验的人来说，这都是很好的读本。希望本书对更多的人有益，也希望中国的“人工智能+”蓬勃发展。

新浪微博算法经理　陶辉

试读结束：需要全本请在线购买：　www.ertongbook.com

前言

起源

在读研究生期间，我就对机器学习算法萌生了很浓的兴趣，并对机器学习中的常用算法进行了学习，利用 MATLAB 对每一个算法进行了实践。在此过程中，每当遇到不懂的概念或者算法时，就会在网上查找相关的资料。也看到很多人在博客中分享算法的学习心得及算法的具体过程，其中有不少内容让我受益匪浅，但是有的内容仅仅是算法的描述，缺少实践的具体过程。

注意到这一点之后，我决定开始在博客中分享自己学习每一个机器学习算法的点点滴滴，为了让更多的初学者能够理解算法的具体过程并从中受益，我计划从三个方面出发，第一是算法过程的简单描述，第二是算法理论的详细推导，第三是算法的具体实践。2014 年 1 月 10 日，我在 CSDN 上写下了第一篇博客。当时涉及的方向主要是优化算法和简单易学的机器学习算法。

随着学习的深入，博客的内容越来越多，同时，在写作过程中，博客的质量也在慢慢提高，这期间也是机器学习快速发展的阶段，在行业内出现了很多优秀的算法库，如 Java 版本的 weka、Python 版本的 sklearn，以及其他的一些开源程序，通过对这些算法库的学习，我丰富了很多算法的知识，同时，我将学习到的心得记录在简单易学的机器学习算法中。工作之后，越发觉得这些基础知识对于算法的理解很有帮助，积累的这些算法学习材料成了我宝贵的财富。

2016 年，电子工业出版社博文视点的符隆美编辑联系到我，询问我是否有意向将这些博文汇总写一本书。能够写一本书是很多人的梦想，我也不例外。于是在 2016

年 9 月，我开始了对本书的构思，从选择算法开始，选择出使用较多的一些机器学习算法。在选择好算法后，从算法原理和算法实现两个方面对算法进行描述，希望本书能够在内容上既能照顾到初学者，又能使具有一定机器学习基础的读者从中受益。

在写作的过程中，我重新查阅了资料，力求保证知识的准确性，同时，在实践的环节中，我使用了目前比较流行的 Python 语言实现每一个算法，使得读者能够更容易理解算法的过程，在介绍深度学习的部分时，使用到了目前最热门的 TensorFlow 框架。为了帮助读者理解机器学习算法在实际工作中的具体应用，本书专门有一章介绍项目实践的部分，综合前面各种机器学习算法，介绍每一类算法在实际工作中的具体应用。

内容组织

本书开篇介绍机器学习的基本概念，包括监督学习、无监督学习和深度学习的基本概念。

第一部分介绍分类算法。分类算法是机器学习中最常用的算法。在分类算法中着重介绍 Logistic 回归、Softmax Regression、Factorization Machine、支持向量机、随机森林和 BP 神经网络等算法。

第二部分介绍回归算法。与分类算法不同的是，在回归算法中其目标值是连续的值，而在分类算法中，其目标值是离散的值。在回归算法中着重介绍线性回归、岭回归和 CART 树回归。

第三部分介绍聚类算法。聚类是将具有某种相同属性的数据聚成一个类别。在聚类算法中着重介绍 K-Means 算法、Mean Shift 算法、DBSCAN 算法和 Label Propagation 算法。

第四部分介绍推荐算法。推荐算法是一类基于具体应用场景的算法的总称。在推荐算法中着重介绍基于协同过滤的推荐、基于矩阵分解的推荐和基于图的推荐。

第五部分介绍深度学习。深度学习是近年来研究最为火热的方向。深度学习的网络模型和算法有很多种，在本书中，主要介绍最基本的两种算法：AutoEncoder 和卷积神经网络。

第六部分介绍以上这些算法在具体项目中的实践。通过具体的例子，可以清晰地看到每一类算法的应用场景。

附录介绍在实践中使用到的 Python 语言、numpy 库及 TensorFlow 框架的具体使用方法。

小结

本书试图从算法原理和实践两个方面来介绍机器学习中的常用算法，对每一类机器学习算法，精心挑选了具有代表性的算法，从理论出发，并配以详细的代码，本书的所有示例代码使用 Python 语言作为开发语言，读者可以从 https://github.com/zhaozhiyong19890102/Python-Machine-Learning-Algorithm 中下载本书的全部示例代码。

由于时间仓促，书中难免存在错误，欢迎广大读者和专家批评、指正，同时，欢迎大家提供意见和反馈。本书作者的电子邮箱：zhaozhiyong1989@126.com。

致谢

首先，我要感谢陶辉和孙永生这两位良师益友，在本书的写作过程中，为我提供了很多意见和建议，包括全书的组织架构。感谢陶辉抽出宝贵的时间帮我写序，感谢孙永生帮我检查程序。

其次，我要感谢符隆美编辑和董雪编辑在写作和审稿的过程中对我的鼓励和悉心指导。

再次，我要感谢姜贵彬、易慧民、潘文彬，感谢他们能够抽出宝贵的时间帮本书写推荐语，感谢他们在读完本书后给出的宝贵意见和建议。

然后，我要感谢 July 在本书的写作过程中对本书提出的宝贵意见，感谢张俊林、王斌在读完本书初稿后对本书的指点。

最后，感谢我的亲人和朋友，是你们的鼓励才使得本书能够顺利完成。

赵志勇

2017 年 6 月 6 日于北京

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，与我们和其他读者一同学习交流。

页面入口：http://www.broadview.com.cn/31319

三步加入“人工智能交流群”，实时获取资源共享，并有机会与大咖实时交流。

- 扫码添加小编为微信好友。
- 申请验证时输入“AI”。
- 小编带你加入“人工智能交流群”。

目录

试读结束：需要全本请在线购买： www.ertongbook.com

第二部分 回归算法

第三部分 聚类算法

试读结束：需要全本请在线购买： www.ertongbook.com