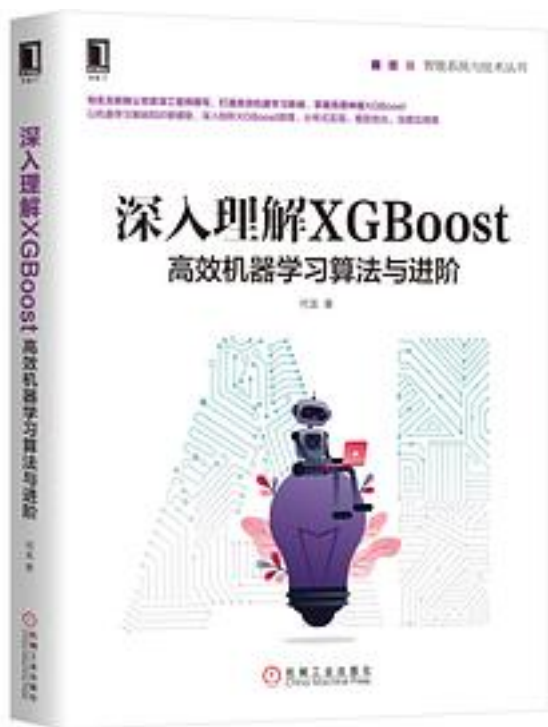


深入理解XGBoost：高效机器学习算法与进阶



[深入理解XGBoost：高效机器学习算法与进阶 下载链接1](#)

著者:何龙

出版者:机械工业出版社

出版时间:2020-1-20

装帧:平装

isbn:9787111642626

本书以机器学习基础知识做铺垫，深入剖析XGBoost的原理、分布式实现、模型优化、深度应用等。

第1~3章使读者对机器学习算法形成整体认知，了解如何优化模型以及评估预测结果，并熟悉常用机器学习算法的实现原理和应用，如线性回归、逻辑回归、决策树、神经网络、支持向量机等。

第4章借助实际案例，讲解如何通过XGBoost解决分类、回归、排序等问题，并介绍了XGBoost常用功能的使用方法。

第5~7章是本书的重点，从理论推导与源码层面深入剖析XGBoost，涵盖XGBoost原理与理论证明、分布式XGBoost的实现、XGBoost各组件的源码解析。

第8~9章为进阶内容，着重解析算法实践与工程应用中的难点，进而帮助读者更好地解决实际问题。

第10章介绍了一些较为前沿的将树模型与其他模型融合的研究方法，以开拓眼界，拓展思路。

作者介绍:

何龙

现就职于滴滴出行，XGBoost开源社区贡献者，专注于人工智能和机器学习领域，从底层算法原理到上层应用实践都有广泛的兴趣和研究。较早接触XGBoost，熟悉XGBoost应用开发，深入阅读源码，具有丰富的项目开发经验。

目录: 前言

第1章 机器学习概述1

1.1 何谓机器学习1

1.1.1 机器学习常用基本概念2

1.1.2 机器学习类型3

1.1.3 机器学习应用开发步骤4

1.2 集成学习发展与XGBoost提出5

1.2.1 集成学习5

1.2.2 XGBoost6

1.3 小结7

第2章 XGBoost骊珠初探9

2.1 搭建Python机器学习环境9

2.1.1 Jupyter Notebook10

2.1.2 NumPy11

2.1.3 Pandas18

2.1.4 Matplotlib32

2.1.5 scikit-learn39

2.2 搭建XGBoost运行环境39

2.3 示例：XGBoost告诉你蘑菇是否有毒42

2.4 小结44

第3章 机器学习算法基础45

3.1 KNN45

3.1.1 KNN关键因素46

3.1.2 用KNN预测鸢尾花品种47

3.2 线性回归52

3.2.1 梯度下降法53

3.2.2 模型评估55

3.2.3 通过线性回归预测波士顿房屋价格55

3.3 逻辑回归57

3.3.1 模型参数估计59

3.3.2 模型评估60

3.3.3 良性/恶性乳腺肿瘤预测61

3.3.4 softmax64

3.4 决策树65

3.4.1 构造决策树66

- 3.4.2 特征选择67
- 3.4.3 决策树剪枝71
- 3.4.4 决策树解决肿瘤分类问题71
- 3.5 正则化75
- 3.6 排序78
 - 3.6.1 排序学习算法80
 - 3.6.2 排序评价指标81
- 3.7 人工神经网络85
 - 3.7.1 感知器85
 - 3.7.2 人工神经网络的实现原理87
 - 3.7.3 神经网络识别手写体数字90
- 3.8 支持向量机92
 - 3.8.1 核函数95
 - 3.8.2 松弛变量97
 - 3.8.3 通过SVM识别手写体数字98
- 3.9 小结99
- 第4章 XGBoost小试牛刀100
 - 4.1 XGBoost实现原理100
 - 4.2 二分类问题101
 - 4.3 多分类问题109
 - 4.4 回归问题113
 - 4.5 排序问题117
 - 4.6 其他常用功能121
 - 4.7 小结145
- 第5章 XGBoost原理与理论证明146
 - 5.1 CART146
 - 5.1.1 CART生成147
 - 5.1.2 剪枝算法150
 - 5.2 Boosting算法思想与实现151
 - 5.2.1 AdaBoost151
 - 5.2.2 Gradient Boosting151
 - 5.2.3 缩减153
 - 5.2.4 Gradient Tree Boosting153
 - 5.3 XGBoost中的Tree Boosting154
 - 5.3.1 模型定义155
 - 5.3.2 XGBoost中的Gradient Tree Boosting156
 - 5.4 切分点查找算法161
 - 5.4.1 精确贪心算法161
 - 5.4.2 基于直方图的近似算法163
 - 5.4.3 快速直方图算法165
 - 5.4.4 加权分位数概要算法167
 - 5.4.5 稀疏感知切分点查找算法167
 - 5.5 排序学习169
 - 5.6 DART174
 - 5.7 树模型的可解释性177
 - 5.7.1 Saabas177
 - 5.7.2 SHAP179
 - 5.8 线性模型原理183
 - 5.8.1 Elastic Net回归183
 - 5.8.2 并行坐标下降法184
 - 5.8.3 XGBoost线性模型的实现185
 - 5.9 系统优化187
 - 5.9.1 基于列存储数据块的并行学习188
 - 5.9.2 缓存感知访问190

5.9.3 外存块计算	191
5.10 小结	192
第6章 分布式XGBoost	193
6.1 分布式机器学习框架Rabit	193
6.1.1 AllReduce	193
6.1.2 Rabbit	195
6.1.3 Rabbit应用	197
6.2 资源管理系统YARN	200
6.2.1 YARN的基本架构	201
6.2.2 YARN的工作流程	202
6.2.3 XGBoost on YARN	203
6.3 可移植分布式XGBoost4J	205
6.4 基于Spark平台的实现	208
6.4.1 Spark架构	208
6.4.2 RDD	210
6.4.3 XGBoost4J-Spark	211
6.5 基于Flink平台的实现	223
6.5.1 Flink原理简介	224
6.5.2 XGBoost4J-Flink	227
6.6 基于GPU加速的实现	229
6.6.1 GPU及其编程语言简介	229
6.6.2 XGBoost GPU加速原理	230
6.6.3 XGBoost GPU应用	236
6.7 小结	239
第7章 XGBoost进阶	240
7.1 模型训练、预测及解析	240
7.1.1 树模型训练	240
7.1.2 线性模型训练	256
7.1.3 模型预测	258
7.1.4 模型解析	261
7.2 树模型更新	264
7.2.1 updater_colmaker	264
7.2.2 updater_histmaker	264
7.2.3 updater_fast_hist	271
7.2.4 其他更新器	276
7.3 目标函数	278
7.3.1 二分类	279
7.3.2 回归	280
7.3.3 多分类	282
7.3.4 排序学习	284
7.4 评估函数	288
7.4.1 概述	289
7.4.2 二分类	291
7.4.3 多分类	295
7.4.4 回归	296
7.4.5 排序	297
7.5 小结	299
第8章 模型选择与优化	300
8.1 偏差与方差	300
8.2 模型选择	303
8.2.1 交叉验证	304
8.2.2 Bootstrap	306
8.3 超参数优化	307
8.3.1 网格搜索	308

8.3.2 随机搜索310
8.3.3 贝叶斯优化313
8.4 XGBoost超参数优化315
8.4.1 XGBoost参数介绍315
8.4.2 XGBoost调参示例319
8.5 小结334
第9章 通过XGBoost实现广告分类器335
9.1 PCA335
9.1.1 PCA的实现原理335
9.1.2 通过PCA对人脸识别数据降维338
9.1.3 利用PCA实现数据可视化341
9.2 通过XGBoost实现广告分类器343
9.3 小结357
第10章 基于树模型的其他研究与应用358
10.1 GBDT、LR融合提升广告点击率358
10.2 mGBDT360
10.3 DEF362
10.4 一种基于树模型的强化学习方法366
10.5 小结370
• • • • • [\(收起\)](#)

[深入理解XGBoost：高效机器学习算法与进阶_下载链接1](#)

标签

机器学习

ml

人工智能

XGBoost

计算机

数据挖掘

数据分析

评论

基本上是一本比较实际的书，整体上来说还是入门水平，比较make sense的地方是数学上并没有讲的太模糊，概念都会有例子从而比较清楚，也有几个项目提供了不少源代码，还没动手但估计自己琢磨pipe花的时间肯定比买JD五折的书的成本高hhh，但是深度上来说确实不足，分布式相关的东西基本上就是浅尝辄止的程度，估计是觉得都是复制粘贴上云的活，但是实际情况也就是这样

一本书讲xgb确实有点夸张了，不过讲的确实不错。看前五章就够了，有机器学习底子的人只看第五章也行。后面都是讲xgb和其他平台的配合使用了。

源码部分写的很深入

很不错的工具集，可以与神经网络相抗衡。终于出了相关书籍了

工作需要，买来看看，内容详实，涉及分布式应用

最近用这个的好多，尤其是做赛题的比较多

[深入理解XGBoost：高效机器学习算法与进阶 下载链接1](#)

书评

[深入理解XGBoost：高效机器学习算法与进阶 下载链接1](#)