

《Python机器学习》

图书基本信息

书名：《Python机器学习》

13位ISBN编号：9787115433739

出版时间：2016-12

作者：[美] Michael Bowles

页数：320

译者：沙赢,李鹏

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《Python机器学习》

内容概要

在学习和研究机器学习的时候，面临令人眼花缭乱的算法，机器学习新手往往会不知所措。本书从算法和Python 语言实现的角度，帮助读者认识机器学习。书专注于两类核心的“算法族”，即惩罚线性回归和集成方法，并通过代码实例来展示所讨论的算法的使用原则。全书共分为7章，详细讨论了预测模型的两类核心算法、预测模型的构建、惩罚线性回归和集成方法的具体应用和实现。本书主要针对想提高机器学习技能的Python 开发人员，帮助他们解决某一特定的项目或是提升相关的技能。

《Python机器学习》

作者简介

Michael Bowles 在硅谷黑客道场教授机器学习，提供机器学习项目咨询，同时参与了多家创业公司，涉及的领域包括生物信息学、金融高频交易等。他在麻省理工学院获得助理教授教职后，创建并运营了两家硅谷创业公司，这两家公司都已成功上市。他在黑客道场的课程往往听者云集并且好评颇多。

书籍目录

第1章关于预测的两类核心算法

- 1.1为什么这两类算法如此有用
- 1.2什么是惩罚回归方法
- 1.3什么是集成方法
- 1.4算法的选择
- 1.5构建预测模型的流程
 - 1.5.1构造一个机器学习问题
 - 1.5.2特征提取和特征工程
 - 1.5.3确定训练后的模型的性能
- 1.6各章内容及其依赖关系
- 1.7小结
- 1.8参考文献

第2章通过理解数据来了解问题

- 2.1“解剖”一个新问题
 - 2.1.1属性和标签的不同类型决定模型的选择
 - 2.1.2新数据集的注意事项
 - 2.2分类问题：用声纳发现未爆炸的水雷
 - 2.2.1“岩石vs水雷”数据集的物理特性
 - 2.2.2“岩石vs水雷”数据集统计特征
 - 2.2.3用分位数图展示异常点
 - 2.2.4类别属性的统计特征
 - 2.2.5利用PythonPandas对“岩石vs水雷”数据集进行统计分析
 - 2.3对“岩石vs水雷数据集”属性的可视化展示
 - 2.3.1利用平行坐标图进行可视化展示
 - 2.3.2属性和标签的关系可视化
 - 2.3.3用热图（heatmap）展示属性和标签的相关性
 - 2.3.4对“岩石vs
 - 2.4基于因素变量的实数值预测鲍鱼的年龄
 - 2.4.1回归问题的平行坐标图—鲍鱼问题的变量关系可视化
 - 2.4.2回归问题如何使用关联热图—鲍鱼问题的属性对关系的可视化
 - 2.5用实数值属性预测实数值目标：评估红酒口感
 - 2.6多类别分类问题：它属于哪种玻璃
- 小结

参考文献

第3章预测模型的构建：平衡性能、复杂性以及大数据

- 3.1基本问题：理解函数逼近
 - 3.1.1使用训练数据
 - 3.1.2评估预测模型的性能
- 3.2影响算法选择及性能的因素——复杂度以及数据
 - 3.2.1简单问题和复杂问题的对比
 - 3.2.2一个简单模型与复杂模型的对比
 - 3.2.3影响预测算法性能的因素
 - 3.2.4选择一个算法：线性或者非线性
- 3.3度量预测模型性能
 - 3.3.1不同类型问题的性能评价指标
 - 3.3.2部署模型的性能模拟
- 3.4模型与数据的均衡

3.4.1通过权衡问题复杂度、模型复杂度以及数据集规模来选择模型

3.4.2使用前向逐步回归来控制过拟合

3.4.3评估并理解你的预测模型

3.4.4通过惩罚回归系数来控制过拟合——岭回归

小结

参考文献

第4章惩罚线性回归模型

4.1为什么惩罚线性回归方法如此有效

4.1.1足够快速地估计系数

4.1.2变量的重要性信息

4.1.3部署时的预测足够快速

4.1.4性能可靠

4.1.5稀疏解

4.1.6问题本身可能需要线性模型

4.1.7什么时候使用集成方法

4.2惩罚线性回归：对线性回归进行正则化以获得最优性能

4.2.1训练线性模型：最小化错误以及更多

4.2.2向OLS公式中添加一个系数惩罚项

4.2.3其他有用的系数惩罚项：Manhattan以及ElasticNet

4.2.4为什么套索惩罚会导致稀疏的系数向量

4.2.5ElasticNet惩罚项包含套索惩罚项以及岭惩罚项

4.3求解惩罚线性回归问题

4.3.1理解最小角度回归与前向逐步回归的关系

4.3.2LARS如何生成数百个不同复杂度的模型

4.3.3从数百个LARS生成结果中选择最佳模型

4.3.4使用Glmnet：非常快速并且通用

4.4基于数值输入的线性回归方法的扩展

4.4.1使用惩罚回归求解分类问题

4.4.2求解超过2种输出的分类问题

4.4.3理解基扩展：使用线性方法来解决非线性问题

4.4.4向线性方法中引入非数值属性

小结

参考文献

第5章使用惩罚线性方法来构建预测模型

5.1惩罚线性回归的Python包

5.2多变量回归：预测红酒口感

5.2.1构建并测试模型以预测红酒口感

5.2.2部署前在整个数据集上进行训练

5.2.3基扩展：基于原始属性扩展新属性来改进性能

5.3二分类：使用惩罚线性回归来检测未爆炸的水雷

5.3.1构建部署用的岩石水雷分类器

5.4多类别分类—分类犯罪现场的玻璃样本

小结

参考文献

第6章集成方法

6.1二元决策树

6.1.1如何利用二元决策树进行预测

6.1.2如何训练一个二元决策树

6.1.3决策树的训练等同于分割点的选择

- 6.1.4二元决策树的过拟合
- 6.1.5针对分类问题和类别特征所做的修改
- 6.2自举集成：Bagging算法
 - 6.2.1Bagging算法是如何工作的
 - 6.2.2Bagging算法小结
- 6.3梯度提升法（GradientBoosting）
 - 6.3.1梯度提升法的基本原理
 - 6.3.2获取梯度提升法的最佳性能
 - 6.3.3针对多变量问题的梯度提升法
 - 6.3.4梯度提升方法的小结
- 6.4随机森林
 - 6.4.1随机森林：Bagging加上随机属性子集
 - 6.4.2随机森林的性能
 - 6.4.3随机森林小结
- 6.5小结
- 6.6参考文献
- 第7章用Python构建集成模型
 - 7.1用Python集成方法工具包解决回归问题
 - 7.1.1构建随机森林模型来预测红酒口感
 - 7.1.2用梯度提升预测红酒品质
 - 7.2用Bagging来预测红酒口感
 - 7.3Python集成方法引入非数值属性
 - 7.3.1对鲍鱼性别属性编码引入Python随机森林回归方法
 - 7.3.2评估性能以及变量编码的重要性
 - 7.3.3在梯度提升回归方法中引入鲍鱼性别属性
 - 7.3.4梯度提升法的性能评价以及变量编码的重要性
 - 7.4用Python集成方法解决二分类问题
 - 7.4.1用Python随机森林方法探测未爆炸的水雷
 - 7.4.2构建随机森林模型探测未爆炸水雷
 - 7.4.3随机森林分类器的性能
 - 7.4.4用Python梯度提升法探测未爆炸水雷
 - 7.4.5梯度提升法分类器的性能
 - 7.5用Python集成方法解决多类别分类问题
 - 7.5.1用随机森林对玻璃进行分类
 - 7.5.2处理类不平衡问题
 - 7.5.3用梯度提升法对玻璃进行分类
 - 7.5.4评估在梯度提升法中使用随机森林基学习器的好处
 - 7.6算法比较
- 小结
- 参考文献

《Python机器学习》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com