[电子书]Machine Learning with Spark PDF下载

本书介绍了用作各种机器学习模型输入的数据集加载和处理的Spark API的基础知识。书中有详细的示例和现实世界的用例，并探索常见的机器学习模型，包括推荐系统，分类，回归，聚类和降维。最后涵盖了一些高级主题，如使用大规模文本数据以及使用Spark Streaming进行在线机器学习和模型评估的方法。

通过本书将学习到以下的知识：

（1）、Create your first Spark program in Scala, Java, and Python

（2）、Set up and configure a development environment for Spark on your own computer, as well as on Amazon EC2

（3）、Access public machine learning datasets and use Spark to load, process, clean, and transform data

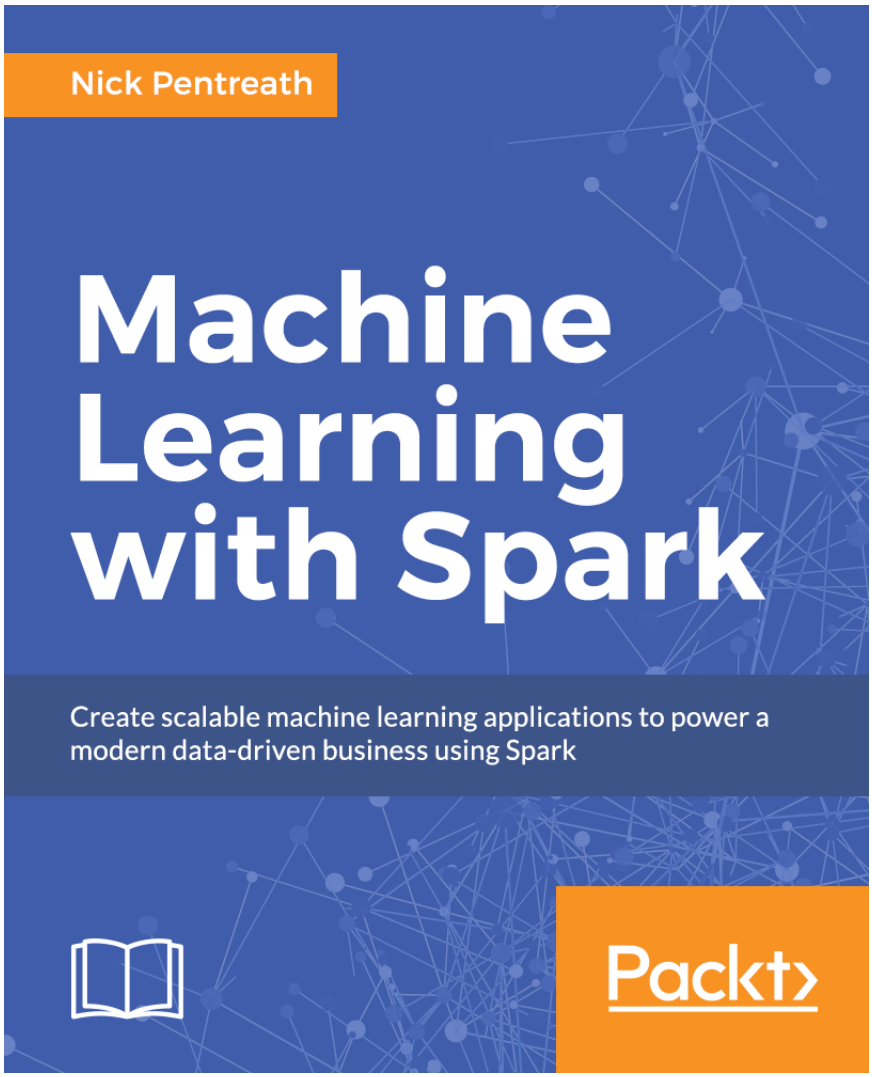
（4）、Use Spark's machine learning library to implement programs utilizing well-known machine learning models including collaborative filtering, classification, regression, clustering, and dimensionality reduction

（5）、Write Spark functions to evaluate the performance of your machine learning models

（6）、Deal with large-scale text data, including feature extraction and using text data as input to your machine learning models

（7）、Explore online learning methods and use Spark Streaming for online learning and model evaluation

Machine Learning with Spark图书由Nick Pentreath所著，全书共338页；Packt Publishing出版社于2015年02月出版。本书作者Nick Pentreath是大数据和机器学习公司Graphflow的联合创始人，该公司专注于推荐和客户智能领域。Nick具有金融市场、机器学习和软件开发的背景。他曾是高盛的研究科学家，定向于伦敦创业认知匹配的在线广告。还层领导过非洲最大的社交网络Mxit的数据科学和分析团队。他热衷于结合商业，使用机器学习和尖端科技打造而成智能系统，通过学习数据带来底线的增值。自2013年以来，Nick一直参与Apache Spark项目，是Apache Spark PMC的成员。



如果想及时了解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

本书的章节

Chapter 1: Getting Up and Running with Spark
Chapter 2: Designing a Machine Learning System
Chapter 3: Obtaining, Processing, and Preparing Data with Spark
Chapter 4: Building a Recommendation Engine with Spark
Chapter 5: Building a Classification Model with Spark
Chapter 6: Building a Regression Model with Spark
Chapter 7: Building a Clustering Model with Spark
Chapter 8: Dimensionality Reduction with Spark
Chapter 9: Advanced Text Processing with Spark
Chapter 10: Real-time Machine Learning with Spark Streaming

下载地址

本文提供了本书的epub、mobi和pdf三种格式下载