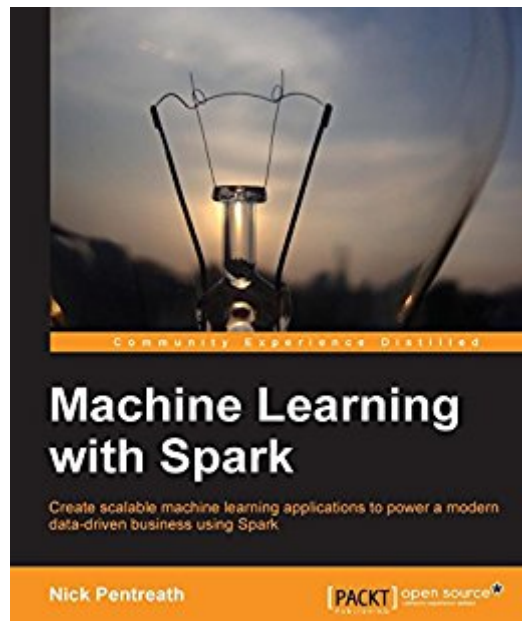The book was found

Machine Learning With Spark – Tackle Big Data With Powerful Spark Machine Learning Algorithms

Community Experience Distilled

Machine Learning with Spark

Create scalable machine learning applications to power a modern data-driven business using Spark

Nick Pentreath

[PACKT] open source

Synopsis

Key FeaturesFollow real-world examples to learn how to develop your own machine learning systems with SparkA practical tutorial with real-world use cases allowing you to develop your own machine learning systems with SparkCombine various techniques and models into an intelligent machine learning systemExplore and use Spark's powerful range of features to load, analyze, clean, and your dataBook DescriptionApache Spark is a framework for distributed computing that is designed from the ground up to be optimized for low latency tasks and in-memory data storage. It is one of the few frameworks for parallel computing that combines speed, scalability, in-memory processing, and fault tolerance with ease of programming and a flexible, expressive, and powerful API design.This book guides you through the basics of Spark's API used to load and process data and prepare the data to use as input to the various machine learning models. There are detailed examples and real-world use cases for you to explore common machine learning models including recommender systems, classification, regression, clustering, and dimensionality reduction. You will cover advanced topics such as working with large-scale text data, and methods for online machine learning and model evaluation using Spark Streaming.What you will learnCreate your first Spark program in Scala, Java, and PythonSet up and configure a development environment for Spark on your own computer, as well as on EC2Access public machine learning datasets and use Spark to load, process, clean, and transform dataUse Spark's machine learning library to implement programs utilizing well-known machine learning models including collaborative filtering, classification, regression, clustering, and dimensionality reductionWrite Spark functions to evaluate the performance of your machine learning modelsDeal with large-scale text data, including feature extraction and using text data as input to your machine learning modelsExplore online learning methods and use Spark Streaming for online learning and model evaluation About the AuthorNick Pentreath is a member of the Apache Spark Project Management Committee. He has has a background in financial markets, machine learning, and software development, including experience as a research scientist at the online ad targeting start-up Cognitive Match Limited in London and leading the Data Science and Analytics team at Mxit, Africa's largest social network. He is also one of the cofounders of Graphflow, a big data and machine learning company focused on user-centric recommendations and customer intelligence.Table of Contents Getting Up and Running with SparkDesigning a Machine Learning SystemObtaining, Processing and Preparing Data with SparkBuilding a Recommendation Engine with SparkBuilding a Classification Model with SparkBuilding a Regression Model with SparkBuilding a Clustering Model with SparkDimensionality Reduction with SparkAdvanced Text Processing with SparkReal-Time Machine Learning with Spark

Streaming

Book Information

File Size: 7907 KB

Print Length: 340 pages

Publisher: Packt Publishing (February 20, 2015)

Publication Date: February 20, 2015

Sold by:Â Digital Services LLC

Language: English

ASIN: B00TXBLFB0

Text-to-Speech: Enabled

X-Ray: Not Enabled

Word Wise: Not Enabled

Lending: Not Enabled

Enhanced Typesetting: Enabled

Best Sellers Rank: #159,921 Paid in Kindle Store (See Top 100 Paid in Kindle Store)   #31 inÂ Books > Computers & Technology > Computer Science > AI & Machine Learning > Machine Theory   #127 inÂ Books > Computers & Technology > Databases & Big Data > Data Processing #1954 inÂ Kindle Store > Kindle eBooks > Computers & Technology

Customer Reviews

This book is a nice introduction to using the Apache Spark framework. It assumes no prior knowledge of either Hadoop, Spark or machine learning itself (although the latter is covered at quite a rapid pace in places so some background would likely be helpful!). The code examples are presented in Python and (mainly) Scala, with examples that are reasonably well-described.The overall tone of the book is clear and the chapters progress in a logical order, with a fairly rapid journey through the main machine learning techniques from a Spark perspective. Later chapters were particularly interesting, covering text mining and more complex methods (e.g. feature hashing).Some of the example data sets feel a little 'tired' (movie ratings data yet again - or perhaps I've just read too many machine learning books), but otherwise this is a good book and comes recommended.

This book gives a great introduction to using the Apache Spark framework, a goto for anyone who wishes to learn how to use Apache Spark framework. It assumes no prior knowledge of either

Hadoop or Spark. The code examples are presented in Python and Scala are well-described.The overall the book is clearly structured and the chapters progress in a logical order, with a fairly rapid introduction to machine learning techniques from a Spark perspective.I definitely do recommend this book for beginners.

This is not a great book.

Machine Learning with Spark - Tackle Big Data with Powerful Spark Machine Learning Algorithms A collection of Advanced Data Science and Machine Learning Interview Questions Solved in Python and Spark (II): Hands-on Big Data and Machine ... Programming Interview Questions) (Volume 7) Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business. Leveraging the Power of Data Analytics, Data ... (Hacking Freedom and Data Driven) (Volume 2) Unsupervised Deep Learning in Python: Master Data Science and Machine Learning with Modern Neural Networks written in Python and Theano (Machine Learning in Python) Deep Learning in Python Prerequisites: Master Data Science and Machine Learning with Linear Regression and Logistic Regression in Python (Machine Learning in Python) Convolutional Neural Networks in Python: Master Data Science and Machine Learning with Modern Deep Learning in Python, Theano, and TensorFlow (Machine Learning in Python) Deep Learning in Python: Master Data Science and Machine Learning with Modern Neural Networks written in Python, Theano, and TensorFlow (Machine Learning in Python) Java Artificial Intelligence: Made Easy, w/ Java Programming; Learn to Create your * Problem Solving * Algorithms! TODAY! w/ Machine Learning & Data ... engineering, r programming, iOS development) Javascript Artificial Intelligence: Made Easy, w/ Essential Programming; Create your * Problem Solving * Algorithms! TODAY! w/ Machine Learning & Data ... engineering, r programming, iOS development) Artificial Intelligence: Made Easy w/ Ruby Programming; Learn to Create your * Problem Solving * Algorithms! TODAY! w/ Machine Learning & Data ... engineering, r programming, iOS development) Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (MIT Press) Java: Artificial Intelligence; Made Easy, w/ Java Programming; Learn to Create your * Problem Solving * Algorithms! TODAY! w/ Machine Learning & Data Structures (Artificial Intelligence Series) Machine Learning: The Art and Science of Algorithms that Make

Sense of Data Javascript Artificial Intelligence: Made Easy, w/ Essential Programming; Create your * Problem Solving * Algorithms! TODAY! w/ Machine Learning & Data Structures (Artificial Intelligence Series) Learning Spark: Lightning-Fast Big Data Analysis Unsupervised Machine Learning in Python: Master Data Science and Machine Learning with Cluster Analysis, Gaussian Mixture Models, and Principal Components Analysis Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Data Structures and Algorithms Made Easy in Java: Data Structure and Algorithmic Puzzles, Second Edition