Yuxi (Hayden) Liu

Python Machine Learning By Example

Easy-to-follow examples that get you up and running with machine learning

Packt>

Python Machine Learning By Example

Data science and machine learning are some of the top buzzwords in the technical world today. The resurgence of interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. This book is your entry point into machine learning.

This book starts with an introduction to machine learning and the Python language, and shows you how to complete the setup. Moving ahead, you will learn all the important concepts, such as exploratory data analysis, data preprocessing, feature extraction, data visualization and clustering, classification, regression, and model performance evaluation. With the help of various projects, you will find it intriguing to acquire the mechanics of several important machine learning algorithms. Also, you will be guided step by step through building your own models from scratch. By the end of the book, you will have developed a broad picture of the machine learning ecosystem and the best practices to use when applying machine learning techniques.

With this book, you will learn to tackle data-driven problems and implement your solutions with the powerful yet simple Python language. The easy-to-follow examples include news topic classification, spam email detection, online ad click-through prediction and stock price forecasting, and will hold your interest through to the end.

Things you will learn:

- Exploit the power of Python to handle data extraction, manipulation, and exploration techniques
- Use Python to visualize data spread across multiple dimensions and extract useful features
- Dive deep into the world of analytics to predict situations correctly
- Implement machine learning classification and regression algorithms from scratch in Python
- Analyze and predict stock market price using the Yahoo/Google Finance data
- Evaluate the performance of a machine learning model and optimize it
- Solve interesting real-world problems using machine learning and Python

Packt>
www.packtpub.com

$ **49.99** US
£ **41.99** UK

Prices do not include local sales
Tax or VAT where applicable

ISBN 978-1-78355-311-2
54999
9 781783 553112

Python Machine Learning By Example

Yuxi (Hayden) Liu

Python Machine Learning By Example

Easy-to-follow examples that get you up and running with machine learning

Yuxi (Hayden) Liu

Packt>

BIRMINGHAM - MUMBAI

Python Machine Learning By Example

Copyright © 2017 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: May 2017

Production reference: 1290517

Published by Packt Publishing Ltd.
Livery Place
35 Livery Street
Birmingham
B3 2PB, UK.

ISBN 978-1-78355-311-2

www.packtpub.com

试读结束：需要全本请在线购买： www.ertongbook.com

Credits

Author
Yuxi (Hayden) Liu

Reviewer
Alberto Boschetti

Commissioning Editor
Veena Pagare

Acquisition Editor
Tushar Gupta

Content Development Editor
Aishwarya Pandere

Technical Editor
Prasad Ramesh

Copy Editor
Safis Editing

Project Coordinator
Nidhi Joshi

Proofreader
Safis Editing

Indexer
Tejal Daruwale Soni

Graphics
Tania Dutta

Production Coordinator
Aparna Bhagat

About the Author

Yuxi (Hayden) Liu is currently a data scientist working on messaging app optimization at a multinational online media corporation in Toronto, Canada. He is focusing on social graph mining, social personalization, user demographics and interests prediction, spam detection, and recommendation systems. He has worked for a few years as a data scientist at several programmatic advertising companies, where he applied his machine learning expertise in ad optimization, click-through rate and conversion rate prediction, and click fraud detection. Yuxi earned his degree from the University of Toronto, and published five IEEE transactions and conference papers during his master's research. He finds it enjoyable to crawl data from websites and derive valuable insights. He is also an investment enthusiast.

About the Reviewer

Alberto Boschetti is a data scientist with strong expertise in signal processing and statistics. He holds a PhD in telecommunication engineering and currently lives and works in London. In his work projects, he faces challenges daily, spanning across natural language processing (NLP), machine learning, and distributed processing. He is very passionate about his job and always tries to be updated on the latest developments of data science technologies, attending meetups, conferences, and other events. He is the author of *Python Data Science Essentials*, *Regression Analysis with Python*, and *Large Scale Machine Learning with Python*, all published by Packt.

I would like to thank my family, my friends, and my colleagues. Also, a big thanks to the open source community.

www.PacktPub.com

For support files and downloads related to your book, please visit `www.PacktPub.com`.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at `www.PacktPub.com` and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at `service@packtpub.com` for more details.

At `www.PacktPub.com`, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.

`https://www.packtpub.com/mapt`

Get the most in-demand software skills with Mapt. Mapt gives you full access to all Packt books and video courses, as well as industry-leading tools to help you plan your personal development and advance your career.

Why subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print, and bookmark content
- On demand and accessible via a web browser

Customer Feedback

Thanks for purchasing this Packt book. At Packt, quality is at the heart of our editorial process. To help us improve, please leave us an honest review on this book's Amazon page at `https://www.amazon.com/dp/1783553111`.

If you'd like to join our team of regular reviewers, you can e-mail us at `customerreviews@packtpub.com`. We award our regular reviewers with free eBooks and videos in exchange for their valuable feedback. Help us be relentless in improving our products!

试读结束：需要全本请在线购买： www.ertongbook.com

Table of Contents

Preface

Data science and machine learning are some of the top buzzwords in the technical world today. A resurging interest in machine learning is due to the same factors that have made data mining and Bayesian analysis more popular than ever. This book is your entry point to machine learning.

What this book covers

`Chapter 1`, *Getting Started with Python and Machine Learning*, is the starting point for someone who is looking forward to enter the field of ML with Python. You will get familiar with the basics of Python and ML in this chapter and set up the software on your machine.

`Chapter 2`, *Exploring the 20 Newsgroups Dataset with Text Analysis Algorithms*, explains important concepts such as getting the data, its features, and pre-processing. It also covers the dimension reduction technique, principal component analysis, and the k-nearest neighbors algorithm.

`Chapter 3`, *Spam Email Detection with Naive Bayes*, covers classification, naive Bayes, and its in-depth implementation, classification performance evaluation, model selection and tuning, and cross-validation. Examples such as spam e-mail detection are demonstrated.

`Chapter 4`, *News Topic Classification with Support Vector Machine*, covers multiclass classification, Support Vector Machine, and how it is applied in topic classification. Other important concepts, such as kernel machine, overfitting, and regularization, are discussed as well.

`Chapter 5`, *Click-Through Prediction with Tree-Based Algorithms*, explains decision trees and random forests in depth over the course of solving an advertising click-through rate problem.

`Chapter 6`, *Click-Through Prediction with Logistic Regression*, explains in depth the logistic regression classifier. Also, concepts such as categorical variable encoding, L1 and L2 regularization, feature selection, online learning, and stochastic gradient descent are detailed.

`Chapter 7`, *Stock Price Prediction with Regression Algorithms*, analyzes predicting stock market prices using Yahoo/Google Finance data and maybe additional data. Also, it covers the challenges in finance and brief explanations of related concepts.

`Chapter 8`, *Best Practices*, aims to foolproof your learning and get you ready for production.

After covering multiple projects in this book, the readers will have gathered a broad picture of the ML ecosystem using Python.

What you need for this book

The following are required for you to utilize this book:

- scikit-learn 0.18.0
- Numpy 1.1
- Matplotlib 1.5.1
- NLTK 3.2.2
- pandas 0.19.2
- GraphViz
- Quandl Python API

You can use a 64-bit architecture, 2GHz CPU, and 8GB RAM to perform all the steps in this book. You will require at least 8GB of hard disk space.

Who this book is for

This book is for anyone interested in entering data science with machine learning. Basic familiarity with Python is assumed.

Conventions

In this book, you will find a number of text styles that distinguish between different kinds of information. Here are some examples of these styles and an explanation of their meaning.

Code words in text, database table names, folder names, filenames, file extensions, pathnames, dummy URLs, user input, and Twitter handles are shown as follows: "The `target_names` key gives the newsgroups names."

试读结束：需要全本请在线购买： www.ertongbook.com