Community Experience Distilled

Building Machine Learning Systems with Python

Master the art of machine learning with Python and build effective machine learning systems with this intensive hands-on guide

Willi Richert
Luis Pedro Coelho

[PACKT] PUBLISHING open source* community experience distilled

Building Machine Learning Systems with Python

Master the art of machine learning with Python and build effective machine learning systems with this intensive hands-on guide

Willi Richert

Luis Pedro Coelho

BIRMINGHAM - MUMBAI

Building Machine Learning Systems with Python

First published: July 2013

Production Reference: 1200713

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

ISBN 978-1-78216-140-0

www.packtpub.com

Cover Image by Asher Wishkerman (a.wishkerman@mpic.de)

Credits

Authors

Willi Richert

Luis Pedro Coelho

Reviewers

Matthieu Brucher

Mike Driscoll

Maurice HT Ling

Acquisition Editor

Kartikey Pandey

Lead Technical Editor

Mayur Hule

Technical Editors

Sharvari H. Baet

Ruchita Bhansali

Athira Laji

Zafeer Rais

Copy Editors

Insiya Morbiwala

Aditya Nair

Alfida Paiva

Laxmi Subramanian

Project Coordinator

Anurag Banerjee

Proofreader

Paul Hindle

Indexer

Tejal R. Soni

Graphics

Abhinash Sahu

Production Coordinator

Aditi Gajjar

Cover Work

Aditi Gajjar

About the Authors

Willi Richert has a PhD in Machine Learning and Robotics, and he currently works for Microsoft in the Core Relevance Team of Bing, where he is involved in a variety of machine learning areas such as active learning and statistical machine translation.

This book would not have been possible without the support of my wife Natalie and my sons Linus and Moritz. I am also especially grateful for the many fruitful discussions with my current and previous managers, Andreas Bode, Clemens Marschner, Hongyan Zhou, and Eric Crestan, as well as my colleagues and friends, Tomasz Marciniak, Cristian Eigel, Oliver Niehoerster, and Philipp Adelt. The interesting ideas are most likely from them; the bugs belong to me.

Luis Pedro Coelho is a Computational Biologist: someone who uses computers as a tool to understand biological systems. Within this large field, Luis works in Bioimage Informatics, which is the application of machine learning techniques to the analysis of images of biological specimens. His main focus is on the processing of large scale image data. With robotic microscopes, it is possible to acquire hundreds of thousands of images in a day, and visual inspection of all the images becomes impossible.

Luis has a PhD from Carnegie Mellon University, which is one of the leading universities in the world in the area of machine learning. He is also the author of several scientific publications.

Luis started developing open source software in 1998 as a way to apply to real code what he was learning in his computer science courses at the Technical University of Lisbon. In 2004, he started developing in Python and has contributed to several open source libraries in this language. He is the lead developer on mahotas, the popular computer vision package for Python, and is the contributor of several machine learning codes.

I thank my wife Rita for all her love and support, and I thank my daughter Anna for being the best thing ever.

About the Reviewers

Matthieu Brucher holds an Engineering degree from the Ecole Superieure d'Electricite (Information, Signals, Measures), France, and has a PhD in Unsupervised Manifold Learning from the Universite de Strasbourg, France. He currently holds an HPC Software Developer position in an oil company and works on next generation reservoir simulation.

Mike Driscoll has been programming in Python since Spring 2006. He enjoys writing about Python on his blog at `http://www.blog.pythonlibrary.org/`. Mike also occasionally writes for the Python Software Foundation, i-Programmer, and Developer Zone. He enjoys photography and reading a good book. Mike has also been a technical reviewer for the following Packt Publishing books: *Python 3 Object Oriented Programming*, *Python 2.6 Graphics Cookbook*, and *Python Web Development Beginner's Guide*.

I would like to thank my wife, Evangeline, for always supporting me. I would also like to thank my friends and family for all that they do to help me. And I would like to thank Jesus Christ for saving me.

Maurice HT Ling completed his PhD. in Bioinformatics and BSc (Hons) in Molecular and Cell Biology at the University of Melbourne. He is currently a research fellow at Nanyang Technological University, Singapore, and an honorary fellow at the University of Melbourne, Australia. He co-edits the Python papers and has co-founded the Python User Group (Singapore), where he has served as vice president since 2010. His research interests lie in life—biological life, artificial life, and artificial intelligence—using computer science and statistics as tools to understand life and its numerous aspects. You can find his website at: `http://maurice.vodien.com`

www.PacktPub.com

Support files, eBooks, discount offers and more

You might want to visit `www.PacktPub.com` for support files and downloads related to your book.

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.

`http://PacktLib.PacktPub.com`

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can access, read and search across Packt's entire library of books.

Why Subscribe?

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

Free Access for Packt account holders

If you have an account with Packt at `www.PacktPub.com`, you can use this to access PacktLib today and view nine entirely free books. Simply use your login credentials for immediate access.

Table of Contents