This book will teach you the fundamental concepts that underpin probability and statistics and illustrates how they relate to machine learning via the Python language and its powerful extensions. This is not a good first book in any of these topics because we assume that you already had a decent undergraduate-level introduction to probability and statistics. Furthermore, we also assume that you have a good grasp of the basic mechanics of the Python language itself. Having said that, this book is appropriate if you have this basic background and want to learn how to use the scientific Python toolchain to investigate these topics. On the other hand, if you are comfortable with Python, perhaps through working in another scientific field, then this book will teach you the fundamentals of probability and statistics and how to use these ideas to interpret machine learning methods. Likewise, if you are a practicing engineer using a commercial package (e.g., Matlab, IDL), then you will learn how to effectively use the scientific Python toolchain by reviewing concepts with which you are already familiar.

The most important feature of this book is that everything in it is reproducible using Python. Specifically, all of the code, all of the figures, and (most of) the text is available in the downloadable supplementary materials that correspond to this book as IPython Notebooks. IPython Notebooks are live interactive documents that allow you to change parameters, recompute plots, and generally tinker with all of the ideas and code in this book. I urge you to download these IPython Notebooks and follow along with the text to experiment with the topics covered. I guarantee doing this will boost your understanding because the IPython Notebooks allow for interactive widgets, animations, and other intuition-building features that help make many of these abstract ideas concrete. As an open-source project, the entire scientific Python toolchain, including the IPython Notebook, is freely available. Having taught this material for many years, I am convinced that the only way to learn is to experiment as you go. The text provides instructions on how to get started installing and configuring your scientific Python environment.

This book is not designed to be exhaustive and reflects the author’s eclectic background in industry. The focus is on fundamentals and intuitions for day-to-day
work, especially when you must explain the results of your methods to a nontechnical audience. We have tried to use the Python language in the most expressive way possible while encouraging good Python coding practices.

Acknowledgments

I would like to acknowledge the help of Brian Granger and Fernando Perez, two of the originators of the Jupyter/IPython Notebook, for all their great work, as well as the Python community as a whole, for all their contributions that made this book possible. Additionally, I would also like to thank Juan Carlos Chavez for his thoughtful review. Hans Petter Langtangen is the author of the Doconce [19] document preparation system that was used to write this text. Thanks to Geoffrey Poore [31] for his work with PythonTeX and LATEX.

San Diego, California
February 2016
Python for Probability, Statistics, and Machine Learning
Unpingco, J.
2016, XV, 276 p. 110 illus., 7 illus. in color., Hardcover
ISBN: 978-3-319-30715-2