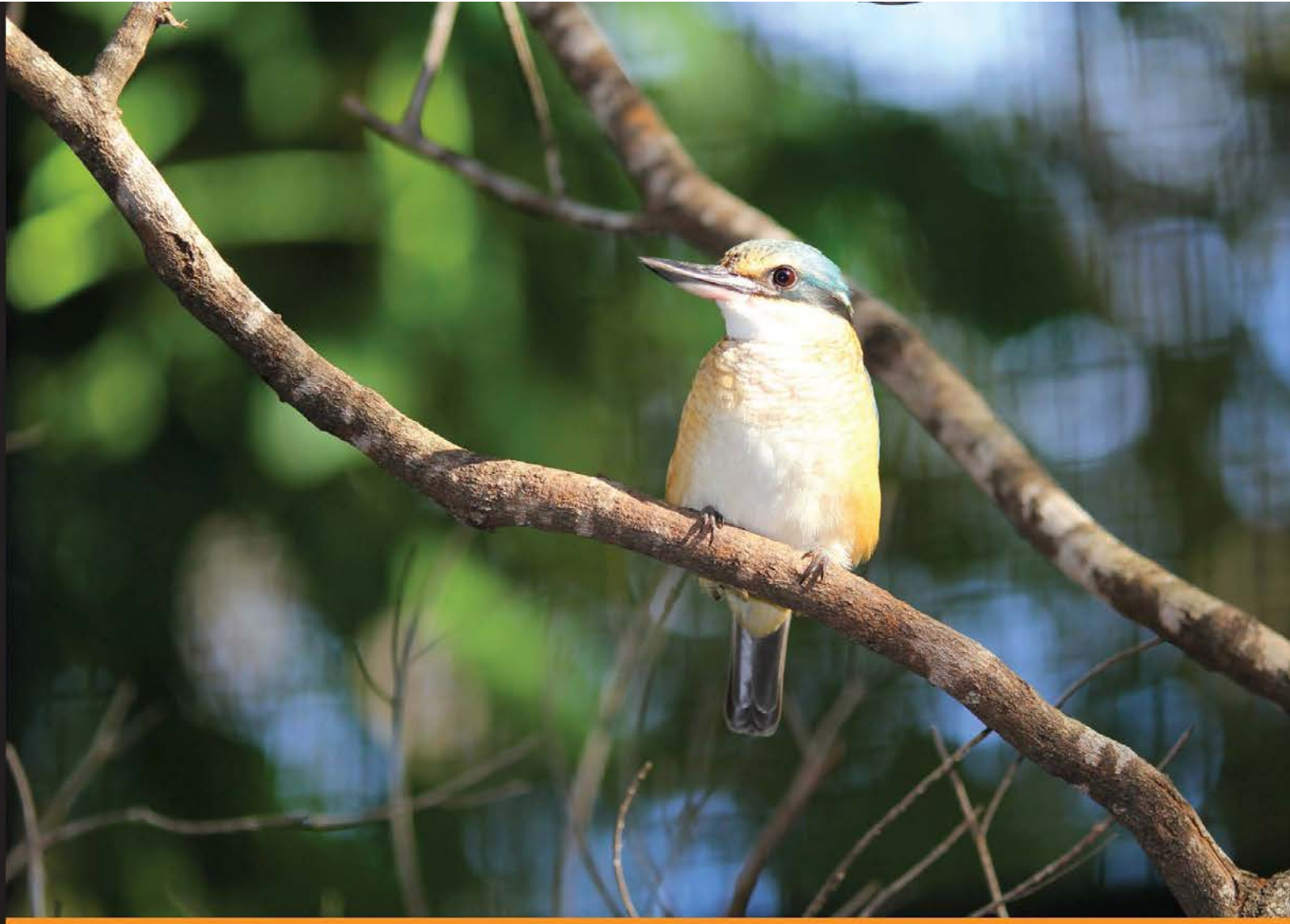

Community Experience Distilled

Learning NumPy Array

Supercharge your scientific Python computations by understanding how to use the NumPy library effectively

Ivan Idris

[PACKT] open source* PUBLISHING
community experience distilled

Learning NumPy Array

Supercharge your scientific Python computations by understanding how to use the NumPy library effectively

Ivan Idris

[PACKT] open source
community experience distilled
PUBLISHING

BIRMINGHAM - MUMBAI

Learning NumPy Array

First published: June 2014

Production Reference: 1060614

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

ISBN 978-1-78398-390-2

www.packtpub.com

Cover Image by Duraid Fatouhi (duraidfatouhi@yahoo.com)

Credits

Author
Ivan Idris

Reviewers
Jonathan Bright
Jaidev Deshpande
Mark Livingstone
Miklós Prisznyák

Commissioning Editor
Kartikey Pandey

Acquisition Editor
Mohammad Rizvi

Content Development Editor
Akshay Nair

Technical Editors
Shubhangi H. Dhamgaye
Shweta S. Pant

Copy Editor
Sarang Chari

Project Coordinator
Lima Danti

Proofreaders
Maria Gould
Kevin McGowen

Indexer
Hemangini Bari

Production Coordinator
Arvindkumar Gupta

Cover Work
Arvindkumar Gupta

About the Author

Ivan Idris has an MSc in Experimental Physics. His graduation thesis had a strong emphasis on applied computer science. After graduating, he worked for several companies as a Java developer, data warehouse developer, and QA analyst. His main professional interests are Business Intelligence, Big Data, and Cloud Computing. He enjoys writing clean, testable code and interesting technical articles. He is the author of *NumPy 1.5 Beginner's Guide* and *NumPy Cookbook*, *Packt Publishing*. You can find more information and a blog with a few NumPy examples at `ivanidris.net`.

I would like to take this opportunity to thank the reviewers and the team at Packt Publishing for making this book possible. Also, I would like to thank my teachers, professors, and colleagues who taught me about science and programming. Last, but not least, I would like to acknowledge my parents, family, and friends for their support.

About the Reviewers

Jonathan Bright has a BS in Electrical Engineering from Rensselaer Polytechnic Institute, and specializes in audio electronics and digital signal processing. He's been programming in Python since import antigravity (the XKCD comic mentioning Python) and contributes to the NumPy and SciPy projects.

Jaidev Deshpande is a software developer at Enthought, Inc., working on software for data analysis and visualization. He's been a research assistant at the University of Pune and Tata Institute of Fundamental Research, working on signal processing and machine learning. He has worked on *Numpy Cookbook*, *Ivan Idris, Packt Publishing*.

Mark Livingstone started his career working for many years for three international computer companies (which no longer exist) in engineering/support/programming/training roles but got tired of being made redundant. He then graduated from Griffith University, Gold Coast, Australia, with a bachelor's degree in Information Technology in 2011. In 2013, he graduated with an honors in B.InfoTech and is currently pursuing his PhD. All his research software is written in Python on a Mac.

Mark enjoys mentoring students with special needs. He is a past chairperson of the IEEE Griffith University Gold Coast Student Branch, volunteers as a qualified Justice of the Peace at the local district courthouse and has been a Credit Union Director. He has also completed 104 blood donations.

In his spare time, he co-develops the Salstat2 statistics package available at `https://sourceforge.net/projects/s2statistical/`, which is multiplatform and uses wxPython, NumPy, SciPy, Scikit, Matplotlib, and a number of other Python modules.

Miklós Prisznyák is a senior software engineer with a scientific background. He graduated as a physicist from the Eötvös Lóránd University, the largest and oldest university in Hungary. He did his MSc thesis on Monte Carlo simulations of non-Abelian lattice quantum field theories in 1992. Having worked for three years in the Central Research Institute for Physics of Hungary, he joined MultiRáció Kft. in Budapest, a company founded by physicists, which specializes in mathematical data analysis and forecasting economic data.

His main project was the Small Area Unemployment Statistics System, which has been in official use at the Hungarian Public Employment Service since then. He learned about the Python programming language there in 2000. He set up his own consulting company in 2002 and then worked on various projects for insurance, pharmacy, and e-commerce companies, using Python whenever he could. He also worked in a European Union research institute in Italy, testing and enhancing a distributed, Python-based Zope/Plone web application.

He moved to Great Britain in 2007 and first worked with a Scottish start-up, using Twisted Python. Then he worked in the aerospace industry in England using, among other things, the PyQt windowing toolkit, the Enthought application framework, and the NumPy and SciPy libraries. He returned to Hungary in 2012 and rejoined MultiRáció, where he's been working on a Python extension module for OpenOffice/EuroOffice, using NumPy and SciPy again, which allows users to solve nonlinear and stochastic optimization and statistical problems.

Miklós likes to travel, read, and he is interested in science, linguistics, history, politics, the board game of Go, and quite a few other topics. Besides these, he always enjoys a good cup of coffee. However, spending time with his brilliant 11-year-old son, Zsombor, is the most important thing for him.

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.

I would like to dedicate this book to the memory of my late uncle, Sahid.
He will be missed.

– Ivan Idris

Table of Contents

Preface **1**

Chapter 1: Getting Started with NumPy **7**

Python **7**
Installing NumPy, Matplotlib, SciPy, and IPython on Windows **8**
Installing NumPy, Matplotlib, SciPy, and IPython on Linux **10**
Installing NumPy, Matplotlib, and SciPy on Mac OS X **11**
Building from source **14**
NumPy arrays **14**
Adding arrays 15
Online resources and help **18**
Summary **18**

Chapter 2: NumPy Basics **19**

The NumPy array object **19**
The advantages of using NumPy arrays 20
Creating a multidimensional array **21**
Selecting array elements **21**
NumPy numerical types **22**
Data type objects 24
Character codes 24
dtype constructors 25
dtype attributes 26
Creating a record data type **26**
One-dimensional slicing and indexing **27**
Manipulating array shapes **28**
Stacking arrays 29
Splitting arrays 33
Array attributes 35
Converting arrays 38

Creating views and copies 39
Fancy indexing 40
Indexing with a list of locations 42
Indexing arrays with Booleans 43
Stride tricks for Sudoku 45
Broadcasting arrays 47
Summary 49

Chapter 3: Basic Data Analysis with NumPy 51

Introducing the dataset 51
Determining the daily temperature range 53
Looking for evidence of global warming 55
Comparing solar radiation versus temperature 57
Analyzing wind direction 61
Analyzing wind speed 62
Analyzing precipitation and sunshine duration 63
Analyzing monthly precipitation in De Bilt 66
Analyzing atmospheric pressure in De Bilt 67
Analyzing atmospheric humidity in De Bilt 69
Summary 71

Chapter 4: Simple Predictive Analytics with NumPy 73

Examining autocorrelation of average temperature with pandas 73
Describing data with pandas DataFrames 76
Correlating weather and stocks with pandas 78
Predicting temperature 79
Autoregressive model with lag 1 79
Autoregressive model with lag 2 80
Analyzing intra-year daily average temperatures 81
Introducing the day-of-the-year temperature model 83
Modeling temperature with the SciPy leastsq function 84
Day-of-year temperature take two 85
Moving-average temperature model with lag 1 87
The Autoregressive Moving Average temperature model 88
The time-dependent temperature mean adjusted autoregressive model 89
Outliers analysis of average De Bilt temperature 92
Using more robust statistics 94
Summary 95

Chapter 5: Signal Processing Techniques **97**
Introducing the Sunspot data **97**
Sifting continued 99
Moving averages **101**
Smoothing functions **103**
Forecasting with an ARMA model **105**
Filtering a signal **107**
Designing the filter 108
Demonstrating cointegration **109**
Summary **112**
Chapter 6: Profiling, Debugging, and Testing **113**
Assert functions **114**
The assert_almost_equal function 114
Approximately equal arrays 115
The assert_array_almost_equal function 116
Profiling a program with IPython **117**
Debugging with IPython **119**
Performing Unit tests **122**
Nose tests decorators **125**
Summary **128**
Chapter 7: The Scientific Python Ecosystem **129**
Numerical integration **129**
Interpolation **130**
Using Cython with NumPy **132**
Clustering stocks with scikit-learn **134**
Detecting corners **137**
Comparing NumPy to Blaze **139**
Summary **140**
Index **141**

Preface

Congratulations on purchasing *Learning NumPy Array*! This was a smart investment, which is guaranteed to save you a lot of time Googling and searching through (online) documentation. You will learn all the essential things needed to become a confident NumPy user. NumPy started originally as part of SciPy and then was singled out as a fundamental library, which other open source Python APIs build on. As such, it is a crucial part of the common Python stack used for numerical and data analysis.

NumPy code is much cleaner than "straight" Python code that tries to accomplish the same task. There are fewer loops required, because operations work directly on arrays and matrices. The many conveniences and mathematical functions make life easier as well. The underlying algorithms have stood the test of time and have been designed with high performance in mind.

NumPy's arrays are stored more efficiently than in an equivalent data structure in base Python, such as in a list of lists. Array IO is significantly faster too. The performance improvement scales with the number of elements of an array. For large arrays, it really pays off to use NumPy. Files as large as several terabytes can be memory-mapped to arrays leading to optimal reading and writing of data. The drawback of NumPy arrays is that they are more specialized than plain lists. Outside of the context of numerical computations, NumPy arrays are less useful.

Large portions of NumPy are written in C. That makes NumPy faster than pure Python code. Finally, since NumPy is open source, you get all of the related advantages. The price is the lowest possible—free as in beer. You don't have to worry about licenses every time somebody joins your team or you need an upgrade of the software. The source code is available to everyone. This of course is beneficial to the code quality.

What this book covers

Chapter 1, *Getting Started with NumPy*, will guide you through the steps needed to install NumPy on your system and helps you create a basic NumPy application. We also successfully run a vector addition program.

Chapter 2, *NumPy Basics*, introduces you to NumPy arrays and fundamentals. In this chapter, we also learn that NumPy arrays can be sliced and indexed in an efficient manner. Here, we will understand the manipulation of shapes of various arrays.

Chapter 3, *Basic Data Analysis with NumPy*, tells us about learning data analysis with weather data analysis as an example. We will also explore the data from a KNMI weather station.

Chapter 4, *Simple Predictive Analytics with NumPy*, helps us attempt to predict the weather with simple models, such as Autoregressive Model with Lag 1 and Autoregressive Model with Lag 2.

Chapter 5, *Signal Processing Techniques*, gives us examples of signal processing and time series analysis. We look at smoothing with window functions and moving averages. We also touch upon the sifting process used by scientists to derive sunspot cycles. And we also get a demonstration of cointegration.

Chapter 6, *Profiling, Debugging, and Testing*, is about profiling, debugging, and testing, which are essential phases in the development cycle. We also cover unit testing, assert functions, and floating-point precision in depth.

Chapter 7, *The Scientific Python Ecosystem*, gives an overview of the Python ecosystem in which NumPy takes a central role. We also examine Cython, which is a relatively young programming language based on Python. We also have a look at Clustering, a type of machine learning algorithm.

What you need for this book

To try out the code samples in this book, you will need a recent build of NumPy. This means that you will need to have one of the Python versions supported by NumPy as well. Some code samples make use of the Matplotlib for illustration purposes. Matplotlib is not strictly required to follow the examples, but it is recommended that you install it too.

Here is a list of software used to develop and test the code examples:

- Python 2.7
- Cython-0.17-py2.7-macosx-10.8-intel.egg