

Numeric PYTHON

Python Data Analysis with NumPy, Pandas,
and Matplotlib

DAS INHALTS- VERZEICHNIS

» Hier geht's
direkt
zum Buch

Contents

Preface	XVII
1 Introduction	1
1.1 The Right Choice	1
1.2 Structure of the Book	2
1.3 This Book and the Tools Behind It	3
1.4 Download the Examples	3
1.5 About the Author	4
1.6 Suggestions and Feedback	4
2 Numerical Programming	5
2.1 Definition of Numerical Programming.....	5
2.2 Overview	5
2.3 The Relationship Between Python, NumPy, Matplotlib, SciPy, and Pandas	6
2.4 Python – An Alternative to MATLAB	7
3 Installation of NumPy, Matplotlib, Pandas, and JupyterLab	9
3.1 Introduction	9
3.2 Installation with conda and Miniconda	10
3.3 Installation with pip.....	11
3.4 Starting JupyterLab	11
3.5 Why JupyterLab?	12

Part I	NumPy	13
4	NumPy Introduction	15
4.1	Overview	15
4.1.1	What is NumPy?	15
4.1.2	A simple example	16
4.2	Comparison of NumPy Data Structures and Lists	17
4.2.1	Key Differences	17
4.2.2	Memory Requirements	18
4.2.3	Time Comparison Between Lists and NumPy Arrays	21
5	Creation and Structure of Arrays	23
5.1	Dimensions	23
5.1.1	Zero-Dimensional Arrays in NumPy	23
5.1.2	One-Dimensional Array	24
5.1.3	Two- and Multi-Dimensional Arrays	24
5.2	Shape of an Array	25
5.3	Indexing and Slicing Operator	26
5.4	Three-Dimensional Arrays	32
5.5	Array Creation Functions	35
5.5.1	arange	35
5.5.2	linspace	37
5.6	Arrays with Zeros and Ones	38
5.7	Identity Matrix	40
5.7.1	The identity Function	40
5.7.2	The eye Function	41
5.8	Data Types	42
5.9	Copying Arrays	44
5.9.1	numpy.copy(A) and A.copy()	44
5.9.2	Contiguous Arrays	44
5.10	Exercises	47
6	Data Type Object: dtype	49
6.1	dtype	49
6.2	Structured Arrays	51

6.3	Input and Output of Structured Arrays	54
6.4	Unicode Strings in Arrays	56
6.5	Renaming Column Names	57
6.6	Replacing Column Values	57
6.7	More Complex Example	58
6.8	Exercises.....	60
7	Combining and Reshaping Arrays.....	61
7.1	Reduction and Reshaping of Arrays.....	61
7.1.1	flatten	62
7.1.2	ravel	62
7.1.3	Differences between ravel and flatten.....	63
7.1.4	reshape	64
7.2	Adding Dimensions	66
7.3	Concatenation and Stacking of Arrays	66
7.3.1	concatenate	67
7.3.2	stack.....	69
7.3.3	dstack.....	72
7.3.4	vstack.....	75
7.3.5	hstack.....	76
7.4	dsplit	78
7.5	Repeating Arrays with tile.....	79
7.6	Exercises.....	82
8	Numerical Operations on NumPy Arrays	83
8.1	Operations with Scalars	83
8.2	Operations between and on Arrays	85
8.3	Matrix Multiplication and Dot Product	86
8.3.1	Definition of the dot Function	86
8.3.2	Examples of the dot Function	87
8.3.3	The dot Product in the Three-Dimensional Case	88
8.4	Comparison Operators	94
8.5	Logical Operators	94
8.6	Broadcasting	95

8.6.1	Row-wise Broadcasting	96
8.6.2	Column-wise Broadcasting	99
8.6.3	Broadcasting with Two One-Dimensional Arrays	102
8.7	Distance Matrix	103
8.8	ufuncs	104
8.8.1	Application of ufuncs	105
8.8.2	Output Parameters in ufuncs	107
8.8.3	accumulate	109
8.8.4	reduce	111
8.8.5	outer	112
8.8.6	at	113
8.9	Exercises.....	113
9	Statistics and Probability	115
9.1	Introduction	115
9.2	Functions Based on the random Module	116
9.2.1	True Random Numbers.....	117
9.2.2	Generating a List of Random Numbers	117
9.2.3	Random Integers	119
9.2.4	Samples or Selections	119
9.2.5	Random Intervals	120
9.2.6	Seed or Initial Value	121
9.2.7	Weighted Random Selection	122
9.2.8	Sampling with Python	125
9.2.9	Cartesian Choice	127
9.2.10	Cartesian Product	127
9.2.11	Cartesian Choice: cartesian_choice	127
9.2.12	Gaussian Normal Distribution	130
9.2.13	Exercise with Binary Transmitter	133
9.3	The random Submodule of NumPy	136
9.3.1	Randomly generating integers and floats.....	136
9.3.2	numpy.random.choice	138
9.3.3	numpy.random.random_sample.....	140
9.4	Synthetic Sales Figures	141
9.5	Exercises.....	143

10	Boolean Masking and Indexing	145
10.1	Fancy Indexing	147
10.2	Indexing with an Integer Array	148
10.3	nonzero and where	148
10.4	Example Applications with <code>np.where</code>	149
10.5	Exercises.....	151
11	Reading and Writing Data Files	153
11.1	Saving text files with <code>savetxt</code>	154
11.2	Loading text files with <code>loadtxt</code>	155
11.2.1	<code>loadtxt</code> without parameters	155
11.2.2	Custom delimiters	156
11.2.3	Selective column reading	156
11.2.4	Data conversion during import	157
11.3	<code>tofile</code>	159
11.4	<code>fromfile</code>	159
11.5	Recommended methods	161
11.6	Another option: <code>genfromtxt</code>	161
Part II	Matplotlib	163
12	Introduction	165
12.1	A first example	166
12.2	Format parameters of <code>plot</code>	167
12.3	Multiple data series with axis labels	169
13	Object-Oriented Plotting	171
13.1	Creating a Figure and Axes	173
13.2	Axis Labels and Title.....	174
13.3	The Plot Method.....	176
13.4	Axis Ranges	177
13.5	Plotting Multiple Functions	179
13.6	Scatter Plots	181
13.7	Filling Areas	184
13.8	Exercises.....	187

14	Multiple Plots and Dual Axes	189
14.1	Subplots with subplot	190
14.2	Flexible Layouts with GridSpec	197
14.3	Dual Axes	204
14.4	Exercises.....	206
15	Axes and Tick Marks	207
15.1	Axes and Spines	207
15.2	Changing Axis Labels	213
15.3	Adjustment of Tick Labels	214
16	Legends and Annotations	215
16.1	Adding a Legend	215
16.2	Annotations.....	219
16.3	Exercises.....	226
17	Contour Plots	227
17.1	Creating a Meshgrid	228
17.2	Functions on Meshgrids	229
17.3	Contour Without Meshgrid	231
17.4	Adjusting Line Styles and Colors	232
17.5	Filled Contours	234
17.6	Custom Colors	235
17.7	Levels.....	236
17.8	Other Grids	237
	17.8.1 Meshgrid in More Detail	237
	17.8.2 mgrid.....	239
	17.8.3 ogrid	240
17.9	imshow	242
17.10	Exercises.....	243
18	Histograms and Diagrams	245
18.1	Histograms	246
18.2	Column Charts	250
18.3	Bar Charts.....	252

18.4	Grouped Bar Charts	253
18.5	xkcd Mode	256
18.6	Pie Charts	258
18.7	Stacked Charts	259
18.8	Exercises.....	260
Part III Pandas		263
19	Pandas:Series	265
19.1	Basics of the Series data structure.....	266
19.2	Access and indexing	269
19.3	Value manipulation with apply	271
19.4	Series from Dictionaries	272
19.5	NaN – Missing Data.....	273
19.5.1	Checking for missing values	274
19.5.2	Relation between NaN and None	274
19.5.3	Filtering missing data	275
19.5.4	Filling missing data	276
19.5.5	Comparison of different interpolation methods	279
19.6	Exercises.....	280
20	DataFrame	281
20.1	A first example	282
20.2	Relation to Series	283
20.3	Manipulating Column Names.....	284
20.4	DataFrames from Dictionaries	285
20.5	Accessing Columns	288
20.6	Row Selection	288
20.6.1	loc	288
20.6.2	query.....	290
20.7	Modification of DataFrames.....	292
20.7.1	Inserting Columns	293
20.7.2	Replacing Columns.....	297
20.7.3	Replacing Rows	298

20.7.4	Modifying Individual Values with <code>at</code> and <code>iat</code>	298
20.8	Changing the Index	299
20.8.1	Reordering Columns and Index	300
20.8.2	Renaming Columns	302
20.8.3	Using a Column as Index	302
20.9	Sums and Cumulative Sums	303
20.9.1	Empty Columns and Filling Them Later	305
20.10	Sorting	306
20.11	Exercises	308
21	Styling	311
21.1	Introduction	311
21.2	Separating Data and Presentation	312
21.3	The <code>.style</code> Property	312
21.3.1	Basic Formatting with <code>.format</code>	313
21.4	Maximum Values in Rows and Columns	313
21.5	Applying a Color Gradient	315
21.5.1	Applying Bar Charts Inside Cells	316
21.6	Exercises	317
22	File Processing	319
22.1	DSV / CSV Files	319
22.1.1	Reading CSV and DSV Files	320
22.1.2	Writing CSV Files	321
22.1.3	Example with a Non-Standard CSV File	325
22.2	Reading and Writing JSON Files	328
22.3	Reading and Writing Excel Files	328
22.4	Exercises	329
23	Pandas: groupby	331
23.1	Groupby with Series	332
23.2	How groupby Works	334
23.3	GroupBy with DataFrames	335
23.3.1	GroupBy with Function	337
23.3.2	Example with File	340
23.4	Exercises	341

24	Pivot Tables	345
24.1	Pivot Function in Pandas	345
24.2	Pivot Call Without Values for values.....	348
24.3	The Function <code>pivot_table</code> in Pandas	349
24.4	Pivoting on the Titanic Data	350
24.5	Exercises.....	354
25	Handling NaN	355
25.1	'nan' in Python	355
25.2	NaN in Pandas	356
25.2.1	Example with NaNs	359
25.3	Using <code>dropna()</code>	362
25.4	Exercises.....	364
26	Binning	365
26.1	Introduction	365
26.2	Binning with Pandas.....	366
26.2.1	Binning with <code>cut</code>	366
26.2.2	Creating an <code>IntervalIndex</code> object	368
26.2.3	More about <code>pd.cut</code>	369
26.2.4	Memory optimization with <code>Categorical</code>	370
26.2.5	Binning with labels	370
26.3	Exercises.....	371
27	Multi-level Indexing	373
27.1	Introduction	373
27.2	Multi-level indexed Series objects	374
27.3	Multi-level indexing through list multiplication	375
27.4	Other ways of creating indices	376
27.5	Access methods	378
27.6	Three-level indices	381
27.7	Relation to DataFrames	383
27.7.1	Manual approach with <code>pd.concat</code>	383
27.7.2	<code>unstack</code> and <code>stack</code>	384
27.8	Swapping multi-level indices	388
27.9	Exercises.....	389

28	Data Visualization with Pandas	391
28.1	Introduction	391
28.2	Line Charts in Pandas	392
28.2.1	Series	392
28.2.2	DataFrames	394
28.2.3	Secondary Axes (Twin Axes)	397
28.2.4	Multiple Y-Axes	398
28.2.5	Converting String Columns to Floats	400
28.3	Bar Charts in Pandas	401
28.3.1	A Simple Example	401
28.3.2	Bar Chart for Programming Language Usage	402
28.3.3	Coloring a Bar Chart	404
28.4	Pie Charts in Pandas	405
28.4.1	A Simple Example	405
28.5	Area Plot with area	407
28.6	Exercises.....	408
29	Time and Date	409
29.1	Introduction	409
29.2	Python Standard Modules for Time Data	410
29.2.1	The date Class	410
29.2.2	The time Class	412
29.3	The datetime Class	413
29.4	Difference Between Times	415
29.4.1	Converting datetime Objects to Strings	416
29.4.2	Conversion with strftime	416
29.5	Output in Local Language	417
29.6	Creating datetime Objects from Strings	419
30	Time Series	421
30.1	Introduction	421
30.2	Time Series and Python.....	422
30.3	Creating Date Ranges	424
30.4	Date Ranges with Time Components	427
30.5	Exercises.....	428

Part IV Applications	429
31 Image Processing Techniques	431
31.1 Introduction	431
31.2 Loading and Displaying Images	432
31.3 Histograms of Color Values	434
31.4 Image Cropping	436
31.5 Geometric Transformations	436
31.6 Filtering	438
31.7 Lightening and Toning Images	443
31.8 Tiling	451
31.9 Watermarking with <code>np.where</code>	452
31.10 Another Example of Watermarking with <code>np.where</code>	454
31.11 Exercises.....	457
32 Financial Management with Pandas	459
32.1 Budget Book	459
32.1.1 Budget Book with CSV File.....	460
32.1.2 Excel budget book with Chart of Accounts	463
32.1.3 Analysis of the Excel budget book.....	465
32.2 Income and expenditure statement.....	467
32.2.1 Journal File	468
32.2.2 Analysis and Visualization of the Data	469
32.2.3 Tax Totals	474
Part V Solutions to the Exercises	477
33 Solutions to the Exercises	479
33.1 Solutions to Chapter 5 (Creation and Structure of Arrays).....	479
33.2 Solutions to Chapter 6 (Data Type Object: <code>dtype</code>)	481
33.3 Solutions to Chapter 7 (Combining and Reshaping Arrays)	483
33.4 Solutions to Chapter 8 (Numerical Operations on NumPy Arrays)	486
33.5 Solutions to Chapter 9 (Statistics and Probability)	489
33.6 Solutions to Chapter 10 (Boolean Masking and Indexing)	494

33.7	Solutions to Chapter 13 (Object-Oriented Plotting)	496
33.8	Solutions to Chapter 14 (Multiple Plots and Dual Axes).....	499
33.9	Solutions to Chapter 16 (Legends and Annotations)	501
33.10	Solutions to Chapter 17 (Contour Plots)	503
33.11	Solutions to Chapter 18 (Histograms and Diagrams).....	507
33.12	Solutions to Chapter 19 (Pandas:Series)	511
33.13	Solutions to Chapter 20 (DataFrame)	515
33.14	Solutions to Chapter 21 (Styling)	520
33.15	Solutions to Chapter 22 (File Processing)	522
33.16	Solutions to Chapter 23 (Pandas: groupby)	527
33.17	Solutions to Chapter 24 (Pivot Tables)	532
33.18	Solutions to Chapter 25 (Handling NaN).....	533
33.19	Solutions to Chapter 26 (Binning)	534
33.20	Solutions to Chapter 27 (Multi-level Indexing)	535
33.21	Solutions to Chapter 28 (Data Visualization with Pandas)	540
33.22	Solutions to Chapter 30 (Time Series)	542
33.23	Solutions to Chapter 31 (Image Processing Techniques)	543
Index	545