

SMM692

Introduction to Programming in Python

Contents

Preface	9
1 Organization of the Module	11
2 Getting Started with Python	13
2.1 Installing Python	13
2.2 How Python Runs Programs	13
2.3 How We Run Python Programs	13
2.4 Managing Python Environments	13
3 Python Language Fundamentals	15
3.1 Variables	15
3.2 Objects	15
3.3 References	15
4 Python Objects	17
4.1 Number Type Fundamentals	18
4.2 String Type Fundamentals	25
4.3 List and Dictionaries	33
4.4 Dictionaries	37
4.5 Tuples	40
4.6 Sets	42
4.7 Files	44
4.8 Python Statements and Syntax	46
4.9 Control Flow (or If-Then Statements)	49
4.10 While and For Loops	51
4.11 Iterations and Comprehensions	55
5 Technical & Scientific Computation with NumPy and SciPy	59
5.1 Installing NumPy and SciPy	59
5.2 NumPy <code>ndarray</code>	59
5.3 Array Creation and Manipulation Methods	65
5.4 Universal Functions in NumPy	80
5.5 Mathematical Methods	80
5.6 Statistical Methods	80
5.7 Linear Algebra	80
5.8 Pseudorandom Number Generation	80
5.9 File Input and Output with <code>ndarrays</code>	80
5.10	80
6 Data Management with Pandas	81
7 Coda	83

Appendices

Appendix A	Cheat Sheets	87
Appendix B	Collaborative and Versioning Tools	95

List of Figures

List of Tables

4.1	Built-In Objects in Python	18
4.2	Number Type Objects in Python	19
4.3	A Sample of Functions Provided by the <code>math</code> Module	20
4.4	Operator Precedence Hierarchy (Ascending Order)	22
4.5	Number Formatting Options in Python	24
4.6	Sample of String Literals and Operators	29
4.7	Popular list methods	35
4.8	Popular dictionary methods	39
4.9	Popular file methods	47
4.10	Python Statements	48
5.1	Common Use Attributes of NumPy <code>array</code>	62
5.2	NumPy Data Types	64
5.3	Routines for Creating Arrays from Shape or Value	67
5.4	Routines for Creating Arrays from Existing Data	69
5.5	Routines for Creating Record Arrays	71
5.6	Routines for Creating Character Arrays	73
5.7	Routines for Numerical Ranges	75
5.8	Routines for Building Matrices	77
5.9	Routines for the Matrix Class	79
A.1	Helpful Escapes	87
A.2	Comprehensive List of String Methods	88
A.3	NumPy Universal Functions: Mathematical Operations	90
A.4	NumPy Universal Functions: Trigonometric Operations	91
A.5	NumPy Universal Functions: Floating Operations	92
A.6	NumPy Universal Functions: Comparison Operations	93

Preface

...

Chapter 1

Organization of the Module

...

Chapter 2

Getting Started with Python

2.1 Installing Python

...

2.2 How Python Runs Programs

...

2.3 How We Run Python Programs

...

2.4 Managing Python Environments

...

Chapter 3

Python Language Fundamentals

3.1 Variables

...

3.2 Objects

...

3.3 References

...

Chapter 4

Python Objects

In this chapter, we pursue the following learning objectives:

-

What is a Python object?

In essence, Python objects are pieces of data. Mark Lutz, the author of the popular book [Learning Python](#)¹, points out

*in Python we do things with stuff.
“Things” take the form of operations like
addition and concatenation, and “stuff”
refers to the objects on which we perform
those operations.*

Built-in and ad-hoc objects

In Python, there are two families of objects: built-in objects provided by the Python language itself and ad-hoc objects — called [classes](#) — we can create to accomplish specific goals.

Why do built-in Python objects matter?

Typically, we do not need to create ad-hoc objects. Python provides us with diverse built-in objects that make our job easier:

- built-in objects make coding efficient and easy. For example, using the [string](#) object, we can represent and manipulate a piece of text — e.g., a newspaper article — without loading any [module](#)
- built-in objects are flexible. For example, we can deploy built-in objects to create a [class](#)
- built-in objects have been created and refined over time by a large community of expert developers. Hence, they are often more efficient than ad-hoc objects (unless the creator of the ad-hoc object really knows her business!)

The core built-in Python objects

Table 4.1 illustrates the types of built-in Python objects. For example, **Numbers** and **strings** objects are used to represent numeric and textual data respectively. **Lists** and **dictionaries** are — likely as not — the two most popular **data structures** in Python. Lists are ordered collections of other objects such (any type!!). Dictionaries are pairs of keys (e.g., a product identifier) and objects (e.g., the price of the product). No worries: we will go through each built-in type in the following sections of this document. Caveat: in the interest of logical coherence, the various built-in types will not be presented in the order adopted Table 4.1.

TABLE 4.1
Built-In Objects in Python

Object type	Example literals/creation
Numbers	1234, 3.1415, 3+4j, 0b111, Decimal(), Fraction()
Strings	'spam', "Bob's", b'a\x01c', u'sp\xc4m'
Lists	[1, [2, 'three'], 4.5], list(range(10))
Dictionaries	{'food': 'spam', 'taste': 'yum'}, dict(hours=10)
Tuples	(1, 'spam', 4, 'U'), tuple('spam'), namedtuple
Files	open('eggs.txt'), open(r'C:\ham.bin', 'wb')
Sets	set('abc'), {'a', 'b', 'c'}
Other core types	Booleans, types, None
Program unit types	Functions, modules, classes
Implementation types	Compiled code, stack tracebacks

4.1 Number Type Fundamentals

Types of 'number' objects

Snippet 4.1, "Doing stuff with numbers," highlights the two most popular **'number'** instances in Python: integers and floating-point numbers. Integers are whole numbers such as 0, 4, or -12. Floating-point numbers are the representation of real numbers such as 0.5, 3.1415, or -1.6e-19. However, floating points in Python do not have — in general — the same value as the real number they represent.² It is worth noticing that any single number with a period '.' is considered a floating point in Python. Also, Snippet 4.1 shows that the multiplication of an integer by a floating point yields a floating point. That happens because Python first converts operands up to the type of the most complicated operand.

Snippet 4.1 — doing 'stuff' with numbers

```
1 # integer addition
2 >>> 1 + 1
3 2
4
5 # floating-point multiplication
6 >>> 10 * 0.5
7 5.0
8
9 # 3 to the power 100
10 >>> 3 ** 100
11 515377520732011331036461129765621272702107522001
```

Besides integers and floating points

Besides integers and floating points numbers, Python includes fixed-precision, rational numbers, Booleans, and sets instances — see Table 4.2.

TABLE 4.2
Number Type Objects in Python

Literal	Interpretation
1234, -24, 0, 999999999999999	Integers (unlimited size)
1.23, 1., 3.14e-10, 4E210, 4.0e+210	Floating-point numbers
0o177, 0x9ff, 0b101010	Octal, hex, and binary literals in 3.X
0177, 0o177, 0x9ff, 0b101010	Octal, octal, hex, and binary literals in 2.X
3+4j, 3.0+4.0j, 3J	Complex number literals
set('spam'), {1, 2, 3, 4}	Sets: 2.X and 3.X construction forms
Decimal('1.0'), Fraction(1, 3)	Decimal and fraction extension types
bool(X), True, False	Boolean type and constants

Basic arithmetic operations in Python

Numbers in Python support the usual mathematical operations:

- '+' addition
- '-' → subtraction
- '*' → multiplication
- '/' → floating point division
- '//' → integer division
- '%' → modulus (remainder)
- '**' → exponentiation

To use these operations, it is sufficient to launch a Python or IPython session without any modules loaded (see Snippet 4.1).

Advanced mathematical operations

Besides the mathematical operations shown above, there are many [modules shipped with Python](#) that carry out advanced/specific numerical analysis. For example, the `math` module provides access to the mathematical functions defined by the [C standard](#).³ Table 4.3 reports a sample of these functions. To use them `math`, we have to import the module as shown in Snippet 4.2. Another popular module shipped with Python is `random`, implementing pseudo-random number generators for various distributions (see the lower section of Example 2).

TABLE 4.3
A Sample of Functions Provided by the `math` Module

Function name	Expression
<code>math.sqrt(x)</code>	\sqrt{x}
<code>math.exp(x)</code>	e^x
<code>math.log(x)</code>	$\ln x$
<code>math.log(x, b)</code>	$\log_b(x)$
<code>math.log10(x)</code>	$\log_{10}(x)$
<code>math.sin(x)</code>	$\sin(x)$
<code>math.cos(x)</code>	$\cos(x)$
<code>math.tan(x)</code>	$\tan(x)$
<code>math.asin(x)</code>	$\arcsin(x)$
<code>math.acos(x)</code>	$\arccos(x)$
<code>math.atan(x)</code>	$\arctan(x)$
<code>math.sinh(x)</code>	$\sinh(x)$
<code>math.cosh(x)</code>	$\cosh(x)$
<code>math.tanh(x)</code>	$\tanh(x)$
<code>math.asinh(x)</code>	$\operatorname{arsinh}(x)$
<code>math.acosh(x)</code>	$\operatorname{arcosh}(x)$
<code>math.atanh(x)</code>	$\operatorname{artanh}(x)$
<code>math.hypot(x, y)</code>	The Euclidean norm, $\sqrt{x^2 + y^2}$
<code>math.factorial(x)</code>	$x!$
<code>math.erf(x)</code>	The error function at x
<code>math.gamma(x)</code>	The gamma function at x , $\Gamma(x)$
<code>math.degrees(x)</code>	Converts x from radians to degrees
<code>math.radians(x)</code>	Converts x from degrees to radians

Snippet 4.2 — advanced mathematical operations with the modules shipped with Python

```
1 # import the math module
2 >>> import math
3
4 # base-y log of x
5 >>> math.log(12, 8)
6 1.1949875002403856
7
8 # base-10 log of x
9 >>> math.log10(12)
10 1.0791812460476249
11
12 # import the random module
13 >>> import random
14
15 # a draw from a normal distribution with mean = 0 and standard deviation = 1
16 >>> random.normalvariate(0, 1)
17 -0.136017752991189
18
19 # trigonometric functions
20 >>> math.cos(0)
21 1.0
22
23 >>> math.sin(0)
24 0.0
25
26 >>> math.tan(0)
27 0.0
28
29 # an expression containing a factorial product
30 >>> math.factorial(4) - 4 * 3 * 2 * 1
31 0
```

Operator precedence

As shown in Snippet 4.2, line 30, Python expressions can string together multiple operators. So, how does Python know which operation to perform first? The answer to this question lies in operator precedence. When you write an expression with more than one operator, Python groups its parts according to what are called precedence rules,⁴ and this grouping determines the order in which the expression's parts are computed. Table 4.4 reports the precedence hierarchy concerning the most common operators. Note that operators lower in the table have higher precedence. Parentheses can be used to create sub-expressions that override operator precedence rules.

TABLE 4.4
Operator Precedence Hierarchy
(Ascending Order)

Operator	Description
<code>x + y</code>	Addition, concatenation
<code>x - y</code>	Subtraction, set difference
<code>x * y</code>	Multiplication, repetition
<code>x % y</code>	Remainder, format;
<code>x / y, x // y</code>	Division: true and floor
<code>-x, +x</code>	Negation, identity
<code>~x</code>	Bitwise NOT (inversion)
<code>x ** y</code>	Power (exponentiation)

Technical and scientific
computation with Python

Python is at the center of a rich ecosystem of modules for technical and scientific computation. In the following chapter, the attention will revolve around two of the most prominent modules: [NumPy](#) and [SciPy](#). In a nutshell, [NumPy](#) offers the infrastructure for the efficient manipulation of (potentially massive) data structures, while [SciPy](#) implements many algorithms across the fields of statistics, linear algebra, optimization, calculus, signal processing, image processing, and others. Another core module in the technical and scientific domain is [SymPy](#), a library for symbolic mathematics. Note that none of these three modules are shipped with Python and should be installed with the package manager of your choice (e.g., [conda](#)).

Variables and Basic Expressions

Variables are simply names—created by you or Python—that are used to keep track of information in your program. In Python:

- Variables are created when they are first assigned values
- Variables are replaced with their values when used in expressions
- Variables must be assigned before they can be used in expressions
- Variables refer to objects and are never declared ahead of time

As Snippet 4.3 shows, the assignment of `x = 2` causes the variable `x` to come into existence ‘automatically.’ From that point, we can use the variables in the context of expressions such as the ones displayed in lines 8, 12, 16, and 20, or to create new variables like in line 24.

Snippet 4.3 — expressions involving arithmetic operations

```
1
2 # let us assign the variables 'x' and 'y' to two number objects
3 >>> x = 2
4
5 >>> y = 4.0
6
7 # subtracting an integer from variable 'x'
8 >>> x - 1
9 1
10
11 # dividing the variable 'y' by an integer
12 >>> y / 73
13 0.0547945205479452
14
15 # integer-dividing the variable 'y' by an integer
16 >>> y // 73
17 0.0
18
19 # getting a linear combination of 'x' and 'y'
20 >>> 3 * x - 5 * y
21 -14.0
22
23 # assigning the variable 'z' to the linear combination of 'x' and 'y'
24 >>> z = 3 * x - 5 * y
```

Displaying number objects

Snippet 4.3 includes some expressions whose result is not passed to a new variable (e.g., lines 8, 12, 16, 20). In those cases, the IPython session displays the outcome of the expression ‘as is’ (e.g., 0.0547945205479452). However, a number with more than three or four decimals may not suit the table or report we have to prepare. Python has powerful [string formatting](#) capabilities to display number objects in a readable and nice manner. Table 4.5 illustrates various number formatting options with concrete cases. Format strings contain ‘replacement fields’ surrounded by curly braces {}. Anything that is not contained in braces is considered literal text, which is copied unchanged to the output. Snippet 4.4 presents a fully-fledged number formatting case. First, we assign the variable `a` to a floating-point number (line 2). Then, we pass the formatting option `{:.2f}` over the variable `a` using the Python built-in function `format`.

TABLE 4.5
Number Formatting Options in Python

Number	Format	Output	Description
3.1415926	{:.2f}	3.14	Format float 2 decimal places
3.1415926	{:+.2f}	+3.14	Format float 2 decimal places with sign
-1	{:+.2f}	-1.00	Format float 2 decimal places with sign
2.71828	{:.0f}	3	Format float with no decimal places
5	{:0>2d}	05	Pad number with zeros (left padding, width 2)
5	{:x<4d}	5xxx	Pad number with x's (right padding, width 4)
10	{:x<4d}	10xx	Pad number with x's (right padding, width 4)
1000000	{:,}	1,000,000	Number format with comma separator
0.25	{:.2%}	25.00%	Format percentage
1000000000	{:.2e}	1.00e+09	Exponent notation
13	{:10d}	13	Right aligned (default, width 10)
13	{:<10d}	13	Left aligned (width 10)
13	{:~10d}	13	Center aligned (width 10)

Snippet 4.4 — number formatting in Python

```

1 # assign the variable 'a' to a floating-point number
2 >>> a = 0.67544908755
3
4 # displaying 'a' with the first two decimals only
5 >>> "{:.2f}".format(a)
6 "0.68"
7
8 # displaying 'a' with the first three decimals only
9 >>> "{:.3f}".format(a)
10 "0.675"

```

How do I compare number objects?

Comparisons are used frequently to create control [flows](#), a topic we will discuss later in this chapter. Normal comparisons in Python regard two number objects and return a Boolean result. Chained comparisons concern three or more objects, and, like normal comparisons, yield a Boolean result. Snippet 4.5 provides a sample of normal comparisons (between lines 1 and 15) and chained comparisons (between lines 21 and 30). As evident in the example, comparisons can regard both numbers and variables assigned to numbers. Chained comparisons can take the form of a range test (see line 21), a joined, ‘AND’ test of the truth of multiple expressions (see line 25) or a disjoined, ‘OR’ test of the truth of multiple expressions (see line 29).

Snippet 4.5 — comparing numeric objects

```
1 # less than
2 >>> 3 < 2
3 False
4
5 # greater than or equal
6 >>> 1 <= 2
7 True
8
9 # equal
10 >>> 2 == 2
11 True
12
13 # not equal
14 >>> 4 != 4
15 False
16
17 # range test
18 >>> x = 3
19 >>> y = 5
20 >>> z = 4
21 >>> x < y < z
22 False
23
24 # joined test
25 >>> x < y and y > z
26 True
27
28 # disjointed test
29 >>> x < y or y < z
30 True
```

4.2 String Type Fundamentals

What is a string?

A Python string is a positionally ordered collection of other objects. Sequences maintain a left-to-right order among the items they contain: their items are stored and fetched by their relative positions. Strictly speaking, strings are *immutable sequences* of one-character strings; other, more general sequence types include lists and tuples, covered later.

How do we use strings?

Strings are used to record words, contents of text files loaded into memory, Internet addresses, Python source code, and so on. Strings can also be used to hold the raw bytes used for media files and network transfers, and both the encoded and decoded forms of non-ASCII Unicode text used in internationalized programs.

Is `abc` a Python string?

String indexing and slicing

Nope. Python strings are enclosed in single quotes (`'...'`) *or* double quotes (`"..."`) with the same result. Hence, `"abc"` can be Python string, while `abc` cannot. `abc` can be a variable name, though.

The fact that strings are immutable sequences affects how we manipulate textual data in Python. In Snippet 4.6, we fetch the individual elements of `S`, a variable assigned to `"Python 3.X."` As per the built-in function `len`, `S` contains six unitary strings. That means that each element in `S` is associated with a position in the numerical progression $\{0, 1, 2, 3, 4, 5\}$. Now, you may be surprised to see the first element of the list is 0 instead of 1. The reason is that Python is a zero-based indexed programming language: the first element of a series has index 0, while the last element has index `len(obj) - 1`. Fetching the individual elements of a string, such as `S`, requires passing the desired index between brackets, as shown in line 9 (where we get the first unitary string, namely, `"P"`), line 13 (where we get the last unitary string, namely, `"X"`), and line 21 (where we get the unitary string with index 3, i.e., the fourth unitary string appearing in `S`, `"h"`). Note that line 17 is an alternative indexing strategy to the one presented in line 13: it is possible to retrieve the last unitary string by counting 'backward'; that is, getting the first element starting from the right-hand side of the string, which equates to index `-1`. In lines 26 and 30, we exploit the indices of `S` to retrieve multiple unitary strings in a row. What we pass among brackets is not a single index. Instead, we specify a range of indices `i:j`. It is worth noticing that, in Python, the element associated with the lower bound index `i` is returned, whereas the element associated with the upper bound index `j` is not. In line 26, we fetch the unitary strings between index 2 — equating to third unitary string of `S` — and index 5 excluded — namely, the fifth unitary string of `S`. In line 30, we adopt the 'backward' approach to retrieve the unitary string with index `-3` — the third string counting from the right-hand side of `S` — as well as any other unitary strings following index `-3`. To do that, we leave the upper bound index blank.

Snippet 4.6 — Python strings as sequences

```
1 # let us assign the string "Python 3.X" to the variable S
2 >>> S = "Python 3.X"
3
4 # check the length of S
5 >>> len(S)
6 6
7
8 # access the first unitary string in the sequence behind S
9 >>> S[0]
10 "P"
11
12 # access the last unitary string in the sequence behind S
13 >>> S[len(S)-1]
14 "X"
15
16 # or, equivalently
17 >>> S[-1]
18 "X"
19
20 # access the i-th, e.g., 3rd, unitary string in the sequence behind S
21 >>> S[3]
22 "h"
23
24 # access the unitary strings between the i-th and j-th positions in the
25 # sequence behind S
26 >>> S[2:5]
27 "tho"
28
29 # access the unitary strings following the i-th position in the sequence
30 # behind S
31 >>> S[-3:]
32 "3.X"
```

Common string literals and operators

Snippet 4.6 deals with string indexing and slicing, two of the many operations we can carry out on strings. Table 4.6 reports a sample of common string literals and operators. The first two lines of Table 4.6 remind us that single and double quotes are equivalent when it comes to assigning a variable to a string object. However, we must refrain from mixing and matching single and double quotes. In other words, a string object requires the leading and trailing quotes are of the same type (i.e., double-double or single-single). In the interest of consistency, it is a good idea to make a policy choice, such as “in my Python code, I use double quotes only”, and to stick with that throughout the various lines of the script. I prefer using double quotes because the single quote symbol is relatively popular in natural language (consider for example the Saxon genitive). As shown in the third line of Table 4.6 the single quote is treated as a unitary string insofar as double quotes are used to delimit the string object. Should the string object be delimited by single quotes, we should tell Python not to treat the single quote symbol after `m` as a Python special character but as a unitary string. To do that, we use the escape symbol `\` as shown in the fourth line of Table 4.6. Table A.1 provides several escaping examples.

TABLE 4.6
Sample of String Literals and Operators

Literal/operation	Interpretation
<code>S = ""</code>	Empty string
<code>S = ''</code>	Single quotes, same as double quotes
<code>S = "spam's"</code>	Single quote as a string
<code>S = 'spam\'s'</code>	Escape symbol
<code>length(S)</code>	Length
<code>S[i]</code>	Index
<code>S[i:j]</code>	Slice
<code>S1 + S2</code>	Concatenate
<code>S * 3</code>	Repeat S <i>n</i> times (e.g., three times)
<code>"text".join(strlist)</code>	Join multiple strings on a character (e.g., "text")
<code>"{}".format()</code>	String formatting expression
<code>S.strip()</code>	Remove white spaces
<code>S.replace("pa", "xx")</code>	Replacement
<code>S.split(",")</code>	Split on a character (e.g., ",")
<code>S.lower()</code>	Case conversion — to lower case
<code>S.upper()</code>	Case conversion — to upper case
<code>S.find("text")</code>	Search substring (e.g., "text")
<code>S.isdigit()</code>	Test if the string is a digit
<code>S.endswith("spam")</code>	End test
<code>S.startswith("spam")</code>	Start test
<code>S = """...multiline..."""</code>	Triple-quoted block strings

String manipulation tasks

Snippet 4.7 presents a sample of miscellaneous string manipulation tasks. In lines 6 and 9, we check the length of the variables `S1` and `S2`. In line 13, we display five repetitions of `S1`. In line 17, we use the algebraic operator `+` to concatenate `S1` and `S2`. In line 21, we expand on the previous input by separating `S1` and `S2` by a white-space. In line 25, we carry out the same task as line 17 — however, we rely on the built-in `join` function to join `S1` and `S2` with whitespace. The argument taken by `join` is a Python `list`, the subject of paragraph 5.3. In line 29, `S1` and `S2` are joined with a custom string object, namely, `" Vs. "`. Finally, in line 33, we use the built-in `format` function (see also Snippet 4.4) to display a string object including `S1` and `S2`. For a comprehensive list of string methods, see Table A.2.

Snippet 4.7 — miscellaneous string manipulation tasks

```
1 # let us assign S1 and S2 to two strings
2 >>> S1 = "Python 3.X"
3 >>> S2 = "Julia"
4
5 # check the length of S1 and S2
6 >>> len(S1)
7 10
8
9 >>> len(S2)
10 5
11
12 # display the S1 repeated five times
13 >>> S1 * 5
14 "Python 3.XPython 3.XPython 3.XPython 3.XPython 3.X"
15
16 # display the concatenation of S1 and S2
17 >>> S1 + S2
18 "Python 3.XJulia"
19
20 # display the concatenation of S1, whitespace, and S2
21 >>> S1 + " " + S2
22 "Python 3.X Julia"
23
24 # display the outcome of joining S1 and S2 with a whitespace
25 >>> " ".join([S1, S2])
26 "Python 3.X Julia"
27
28 # display the outcome of joining S1 and S2 with an arbitrary string object
29 >>> " Vs. ".join([S1, S2])
30 "Python 3.X Vs. Julia"
31
32 # string formatting
33 >>> "Both {} and {} have outstanding ML modules".format(S1, S2)
34 "Both Python 3.X and Julia have outstanding ML modules"
```

String editing

Snippet 4.8 illustrates some string editing tasks. In line 5, we use `rstrip` — a variation of the built-in function `strip` — that returns a copy of the string with leading characters removed. In line 9, we use the built-in `replace` to return a copy of the string with all occurrences of substring `old` (first argument taken by the function) replaced by `new` (second argument taken by the function). Finally, in line 17, we use the built-in function `lower` to return a copy of the string with all the cased characters converted to lowercase.

Snippet 4.8 — miscellaneous string editing tasks

```
1 # let us assign S to a string object
2 >>> S = "Both Python 3.X and Julia have outstanding ML modules"
3
4 # strip target leading characters
5 >>> S.rstrip("Both ")
6 "Python 3.X and Julia have outstanding ML modules"
7
8 # replace target characters
9 >>> S.replace("Python 3.X", "R")
10 "Both R and Julia have outstanding ML modules"
11
12 # split string on target characters
13 >>> S.split(" and ")
14 ["Both Python 3.X", "Julia have outstanding ML modules"]
15
16 # make the string lower case
17 >>> S.lower()
18 "both python 3.x and julia have outstanding ml modules"
```

Testing and searching strings

Snippet 4.9 presents a series of string test and search tasks. The built-in function `find` (see lines 5 and 9) returns the lowest index in the string where substring `sub` is found within the slice `S[start:end]` or `-1` if substring is not found. The built-in function `isdigit` return `True` if all characters in the string are digits and there is at least one character, `False` otherwise. Finally, the built-in function `endswith` returns `True` if the string ends with the specified suffix, otherwise returns `False`.

Snippet 4.9 — miscellaneous string test and search tasks

```
1 # let us assign S to a string object
2 >>> S = "The first version of Python was released in 1991"
3
4 # search for "Python" in S
5 >>> S.find("Python")
6 21
7
8 # search for "Julia" in S
9 >>> S.find("Julia")
10 -1
11
12 # slice the string the get Python's release year information
13 >>> SS = S[-4:]
14
15 # display SS
16 >>> SS
17 "1991"
18
19 # test if all characters in SS are digits
20 >>> SS.isdigit()
21 True
22
23 # test if all characters in SS are digits
24 >>> SS.isdigit()
25 True
26
27 # test if all characters in SS are digits
28 >>> SS.isdigit()
29 True
30
31 # test if S ends with "1991" / SS
32 >>> S.endswith(SS)
33 True
```


Multiline string printing

In the previous examples, we came across the built-in function `print`. Such a function can be used to print both number- and string-type objects. Sometimes, what we want to print fits into a single line. In other circumstances, we are interested in visualizing rich data, which can span multiple lines. Snippet 4.9 how to print objects across multiple lines with the triple-quoted block string (see line). As evident from the Python code included in lines 8-13, any line between triple-quotes is considered part of the same string object.

Snippet 4.9 — multiline string printing

```

1 # single-line print
2 >>> print("Hello world!")
3 Hello world!
4
5 # multi-line print
6 >>> print(
7 ... """
8 ... =====
9 ... COL A          | COL B          | ...          | COL K
10 ... -----
11 ... Sheldon       | Cooper       | ...          | bazinga.com
12 ... -----
13 ... NOTES: this table has fake data
14 ... """
15 ... )
16
17 =====
18 COL A          | COL B          | ...          | COL K
19 -----
20 Sheldon       | Cooper       | ...          | bazinga.com
21 -----
22 NOTES: this table has fake data

```

4.3 List and Dictionaries

What is a list?

A Python `list` is an *ordered, mutable* array of objects. A list is constructed by specifying the objects, separated by commas, between square brackets, `[]`.

Why do we use lists?

Lists are just places to collect other objects so you can treat them as groups.

What type of objects can we include in a list?

Lists can contain any sort of object: numbers, strings, and even other lists. See Snippet 4.10.

Snippet 4.10 — sample lists with different items

```
1 # an empty list
2 >>> L = []
3
4 # a list with an integer, a float, and a string
5 >>> L = [2, -3.56, "XyZ"]
6
7 # a list with an integer and a list
8 >>> L = [4, ["abc", 8.98]]
```

List indexing

We can retrieve one or more list component objects via indexing. That is possible because list items are ordered by their position (similarly to strings). Since Python is a zero-based indexed programming language, to fetch the first item of a list we have to call the index 0 (see Snippet 4.11, line 5). A list nested in another list can be fetched by using multiple indices (line 13). The first index refers to the outer list, while any subsequent index refers to an inner list. In our case, we have two indices, one for each list; that is, `L` and its sub-list `["abc", 8.98]`.

Snippet 4.11 — list indexing and slicing

```
1 # the list
2 >>> L = [4, ["abc", 8.98]]
3
4 # get the first item of L
5 >>> L[0]
6 4
7
8 # get the second element of L
9 >>> L[1]
10 ["abc", 8.98]
11
12 # get the first item of L's second item
13 >>> L[1][0]
14 "abc"
```

List mutability

Lists are mutable objects, which may be changed in place by assignment to offsets and slices, list method calls, deletion statements, and more. Snippet 4.12 illustrates some snippets to change a list's items. In the first part of the example, we change the items using indexing (line 5) and slicing (line 10). In the second half, we use Python's `del` statement to delete the items using indexing (line 15) and slicing (line 20).

Snippet 4.12 — changing and deleting list items in place

```

1 # the list
2 >>> L = ["Leonard", "Penny", "Sheldon"]
3
4 # change the second item of L via indexing
5 >>> L[1] = "Raj"
6 >>> print(L)
7 ["Leonard", "Raj", "Sheldon"]
8
9 # change multiple items of L via slicing
10 >>> L[0:2] = ["Amy", "Howard"]
11 >>> print(L)
12 ["Amy", "Howard", "Sheldon"]
13
14 # delete the first item of L via indexing and using the 'del' statement
15 >>> del L[0]
16 >>> print(L)
17 ["Howard", "Sheldon"]
18
19 # delete multiple items of L via slicing and using the 'del' statement
20 >>> del L[0:2]
21 >>> print(L)
22 []

```

List manipulation with built-in methods

Python offers many [methods to manipulate and test list objects](#). Table 4.7 reports some of the most popular methods along with synopses. The first three methods — `.append()`, `.insert()`, and `.extend()` — expand an existing list. The fourth method, `.index` test for the presence of an item in an existing list. Note this method raises a [ValueError](#) if there is no such item. The remaining methods produce in-place changes in an existing list's items.

TABLE 4.7
Popular list methods

Method	Synopsis
<code>L.append(X)</code>	Append an item to an existing list
<code>L.insert(i, X)</code>	Append an item to an existing list in position <i>i</i>
<code>L.extend([X0, X1, X2])</code>	Extend an existing list with the items from another list
<code>L.index(X)</code>	Get the index of the first instance of the argument in an existing list
<code>L.count(X)</code>	Get the cardinality of an item in an existing list
<code>L.sort()</code>	Sort the items in an existing list
<code>L.reverse()</code>	Reverse the order of the items in an existing list
<code>L.copy()</code>	Get a copy of an existing list
<code>L.pop(i)</code>	Remove the item at the given position in the list, and return it
<code>L.remove(X)</code>	Remove the first instance of an item in an existing list
<code>L.clear()</code>	Remove all items in an existing list

Expanding an existing list

Both the `.append()` and `.extend()` methods can be used to expand an existing list as per Snippet 4.13. However, they accomplish different goals and should not be confused: `.append()` adds a new item (of any type) to the end of the list (see line 6); `.extend()` extend the list by appending all the items from another iterable (e.g., another list, see line 11).

Snippet 4.13 — methods for expanding an existing list

```
1 # create two lists
2 >>> L1 = ["Leonard", "Penny", "Sheldon"]
3 >>> L2 = ["Howard", "Raj", "Amy", "Bernadette"]
4
5 # expand an existing list with .append()
6 >>> L2.append("Priya")
7 >>> print(L2)
8 ["Howard", "Raj", "Amy", "Bernadette", "Priya"]
9
10 # concatenate L1 and L2 with .extend()
11 >>> L1.extend(L2)
12 >>> print(L1)
13 ["Leonard", "Penny", "Sheldon", "Howard", "Raj", "Amy", "Bernadette", "Priya"]
```

In-place change of an existing list's items

One of the most common list manipulation task consists of changing the order of an item's list. As shown in Snippet 4.14, it is possible to use the `.reverse()` method to reverse the elements of the list in place (see line 5), while sorting an item's list can be carried out with the `.sort()` method.

Snippet 4.14 — methods for changing list items in place

```
1 # create a list
2 >>> L = ["Howard", "Raj", "Amy", "Bernadette", "Priya"]
3
4 # reverse the list's item positions
5 >>> L.reverse()
6 >>> print(L)
7 ["Priya", "Bernadette", "Amy", "Raj", "Howard"]
8
9 # sort the list's items
10 >>> L.sort()
11 >>> print(L)
12 ["Amy", "Bernadette", "Howard", "Priya", "Raj"]
```

4.4 Dictionaries

What is a dictionary?

Along with lists, dictionaries are one of the most flexible built-in data types in Python. If you think of lists as ordered collections of objects, you can think of dictionaries as unordered collections; the chief distinction is that in dictionaries, items are stored and fetched by *key*, instead of by *positional offset*.

Why do we use dictionaries?

Dictionaries take the place of records, search tables, and any other sort of aggregation where item names are more meaningful than item positions.

What type of objects can we include in a dictionary?

Like lists, dictionaries can contain objects of any type, and they support nesting to any depth (they can contain lists, other dictionaries, and so on). Each key can have just one associated value, but that value can be a collection of multiple objects if needed, and a given value can be stored under any number of keys.

How do we create a dictionary?

Snippet 4.15 shows two different ways to create a dictionary. A dictionary can be created by including key-value pairs among braces (see line 2). In the example, there are three keys, associated with Marvel characters, and as many values, which can be thought as the characters' position in an ideal power rank. A colon separates a key and its associated value. The second way to create a dictionary is based on Python's builtin `dict`, mapping key onto values, and `zip`, which iterates over two elements in parallel. Specifically, `zip` creates the one-to-one correspondence between keys (characters) and values (characters' power) that is passed as the argument of `dict`. We will analyze the topic of iterations extensively in sections 4.10 and 4.11.

Snippet 4.15 — initializing a new dictionary object

```
1 # method 1
2 >>> D = {"Captain Marvel": 3, "Living Tribunal": 2, "One-Above-All": 1}
3
4 # method 2
5 >>> CHARACTERS = ["Captain Marvel", "Living Tribunal", "One-Above-All"]
6 >>> RANK = [3, 2, 1]
7 >>> D = dict(zip(CHARACTERS, RANK))
8 >>> print(D)
9 {"Captain Marvel": 3, "Living Tribunal": 2, "One-Above-All": 1}
```

Accessing a dictionary's values

Dictionaries' items cannot be accessed via positional offsets — like lists. Instead, we fetch the individual items by using the dictionary keys as shown in Snippet 4.16 (see line 5). The reference key is passed among brackets. When the dictionary at hand contains nested dictionaries (see line 9), it is possible to concatenate multiple queries, namely, sequences of keys between brackets (see line 21).

Snippet 4.16 — fetching dictionary items

```
1 # the dictionary
2 >>> D = {"Captain Marvel": 3, "Living Tribunal": 2, "One-Above-All": 1}
3
4 # let's fetch Captain Marvel's position in the Marvel characters' power rank
5 >>> D["Captain Marvel"]
6 3
7
8 # a dictionary of dictionaries
9 >>> D = {
10     "Dr. Strange": {
11         "first_appearance": 1963,
12         "created_by": "Lee & Ditko"
13     },
14     "Iron Man": {
15         "first_appearance": 1963,
16         "created_by": "Lee, Lieber, Heck & Kirby"
17     },
18 }
19
20 # let us fetch the creator of Dr. Strange
21 >>> D["Dr. Strange"]["created_by"]
22 "Lee & Ditko"
```

Are dictionaries mutable?

Dictionaries, like lists, are mutable. Thus, we can change, expand, and shrink them in place without making new dictionaries: simply assign a value to a key to change or create an entry. The `del` statement works here, too; it deletes the entry associated with the key specified as an index (see Snippet 4.17).

Snippet 4.17 — dictionary mutability examples

```

1 # the dictionary
2 >>> D = {"Captain Marvel": 3, "Living Tribunal": 2, "One-Above-All": 1}
3
4 # let us change the power rank for Captain Marvel
5 >>> D["Captain Marvel"] = 12
6 >>> print(D)
7 {"Captain Marvel": 12, "Living Tribunal": 2, "One-Above-All": 1}
8
9 # let us eliminate the character Living Tribunal
10 >>> del D["Living Tribunal"]
11 >>> print(D)
12 {"Captain Marvel": 12, "One-Above-All": 1}
13
14 # let us add a further character
15 >>> D["Wanda Maximoff"] = 4
16 >>> print(D)
17 {"Captain Marvel": 12, "One-Above-All": 1, "Wanda Maximoff": 4}

```

Dictionary manipulation with built-in methods

Like for lists, Python offers many [methods to manipulate dictionary objects](#). Table 4.8 reports some of the most common methods along with synopses. The first three methods, `.keys()`, `.values()`, `.items()`, get the constitutive elements of dictionaries: keys, values, and key-value pairs respectively. The fourth method, `.get(key, default?)` gets the value for a specific key. The fifth method, `.update()`, updates the value for a specific key. Like `.update()`, `.popitem()`, `.pop()`, and `d.clear()` alter the information of a dictionary in place. The first removes the value of a certain key; the second removes the item (a key-value pair) for a certain key; the latter delete all dictionary items. Finally, `.copy()` creates a [shallow copy](#) of an existing dictionary.

TABLE 4.8
Popular dictionary methods

Method	Synopsis
<code>D.keys()</code>	Get all dictionary keys
<code>D.values()</code>	Get all dictionary values
<code>D.items()</code>	Get all dictionary key-value pairs as tuples
<code>D.get(key, default?)</code>	Query a dictionary element by key
<code>D.update(D2)</code>	Update a dictionary key's value
<code>D.popitem()</code>	Remove the value corresponding to a certain key
<code>D.pop(key, default?)</code>	Remove the item at the given position in the list,
<code>D.clear()</code>	Delete all dictionary items
<code>D.copy()</code>	Copy the target dictionary

How do we access the information in a dictionary?

Snippet 4.18 shows how to use builtin methods to carry out three fundamental tasks: accessing dictionary keys (see line 5), values (see line 9), and items (i.e., key-value pairs, see line 13). It is worth noticing that the three methods illustrated in the example yield specific [dictionary objects](#) such as `dict_keys`, `dict_values`, and `dict_items`. Translating one of these dictionary objects into a list — if needed — is straightforward (see line 17).

Snippet 4.18 — accessing the information included in a dictionary

```

1 # the dictionary
2 >>> D = {"Captain Marvel": 3, "Living Tribunal": 2, "One-Above-All": 1}
3
4 # get the keys
5 >>> D.keys()
6 dict_keys(["Captain Marvel", "Living Tribunal", "One-Above-All"])
7
8 # get the values
9 >>> D.values()
10 dict_values([3, 2, 1])
11
12 # get the items
13 >>> D.items()
14 dict_items([("Captain Marvel", 3), ("Living Tribunal", 2), ("One-Above-All", 1)])
15
16 # get the keys as a list
17 >>> list(D.keys())
18 ["Captain Marvel", "Living Tribunal", "One-Above-All"]

```

4.5 Tuples

What is a tuple?

Tuples are sequences of immutable Python objects. They are similar to lists, but they are immutable. Tuples are created by enclosing a comma-separated list of values in parentheses.

Tuples are immutable!

Tuples are immutable, which means that once they are created, they cannot be changed!!

Why do we use tuples?

Tuples are useful for storing data that is not to be changed, such as the coordinates of a point in a two-dimensional space. In general, we use tuples any time information integrity is a concern — in other words when we want to make sure the information included in an object will not change because of another reference somewhere in our program.

How do we create a tuple?

Python objects, separated by a comma, must be included between parentheses (see Snippet 4.19, line 2).

How do we access the information in a tuple?

By positional offsets, like lists (see Snippet 4.19, lines 5 and 9).

Snippet 4.19 — creating and accessing a tuple

```
1 # the tuple
2 >>> T = ("Captain Marvel", 3)
3
4 # access a tuple element
5 >>> T[0]
6 "Captain Marvel"
7
8 # access a tuple element
9 >>> T[1]
10 3
```

Can we convert a tuple into a list?

Yes, we can. To do that, we pass the tuple as the argument of `list` (see Snippet 4.20).

Snippet 4.20 — tuple conversion

```
1 >>> T = ("Captain Marvel", 3)
2
3 # from a tuple to a list
4 >>> L = list(T)
5 >>> print(L)
6 ["Captain Marvel", 3]
7
8 # amend L's items
9 >>> L[1] = 4
10
11 # get back to a tuple
12 >>> T = tuple(L)
13 >>> print(T)
14 ("Captain Marvel", 4)
```

Tuples with the `collections` module

`collections` is a module that is shipped with Python and provides data containers that are alternative to Python's general purpose built-in containers, i.e., `dict`, `list`, `set`, and `tuple`. One of these containers can be created with the function `namedtuple` (see Snippet 4.21), which allows annotating the tuple items with names. In line 2, we import the function `namedtuple` from the `collections` module. In line 5, we create an ad hoc class that best represents the structure of our sample data, concerning Marvel characters' names and the year in which they first appeared in the comic series. The first argument taken by the function is customary and regards the name of the class we are about to create. The second argument is a list with the names of the attributes included in our data structure. In line 8, we use the newly created class `Rec` to create a tuple, which is eventually printed as per line 11.

Snippet 4.21 — creating an annotated tuple with the collection module

```

1 # import the namedtuple function from the module collection
2 >>> from collections import namedtuple
3
4 # create an ad hoc class object 'Rec' that fits our data structure
5 >>> Rec = namedtuple("Rec", ["character", "first_appearance"])
6
7 # use the generated class "Rec"
8 >>> IRONMAN = Rec("Iron Man", 1963)
9
10 # A named-tuple record
11 >>> IRONMAN
12 Rec(character="Iron Man", first_appearance=1963)

```

4.6 Sets

What is a set?

A `set` is an *unordered* collection of *unique* and *immutable* objects.

What does it mean that sets are unordered collections?

By design, `set` is a data structure with *undefined element ordering* (see Snippet 4.22 — the outcome included in line 6 does not follow any particular order).

What does it mean that sets have unique items?

By definition, an item appears only once in a set, no matter how many times it is added (see Snippet 4.22, line 2 Vs. line 7).

Snippet 4.22 — creating a set

```
1 # create a list
2 >>> L = ["a", "a", "b", "c", "c"]
3
4 # get a set from L
5 >>> S = set(L)
6 >>> print(S)
7 {"b", "a", "c"}
```

Why do we use sets?

Sets made this way support common mathematical set operations (see Snippet 4.23). Hence, they have a variety of applications, especially in numeric and database-focused work.

Snippet 4.23 — set operations

```
1 # create two sets
2 >>> X = set(["a", "b", "c"])
3 >>> Y = set(["c", "d", "e"])
4
5 # set difference
6 >>> X - Y
7 set()
8 >>> Y - X
9 {"c", "b"}
10
11 # union
12 >>> X | Y
13 {"a", "b", "c", "d", "e"}
14
15 # intersection
16 >>> X & Y
17 {"c"}
18
19 # superset
20 >>> X > Y
21 False
22
23 # subset
24 >>> X < Y
25 False
```

4.7 Files

How do the files in our OS relate with Python?

How do we open a file?

How do we source the data stored in a file?

Our Python program may involve input and/or output operations. In other words, we may want to read data from a file stored in our machine and/or write the outcome of our analysis to a file. The built-in function `open` creates a Python file object, which serves as a link to a file residing on your machine. As Lutz notes:

“Compared to the types you’ve seen so far, file objects are somewhat unusual. They are considered a core type because they are created by a built-in function, but they’re not numbers, sequences, or mappings, and they don’t respond to expression operators; they export only methods for common file-processing tasks” (page 282)

We open a pipe to a file using the built-in function `open`. The output of the function is a file object.

We open a pipe to a file using the built-in function `open`. The output of the function is a file object. Snippet 4.24 illustrates how to use `open` for data sourcing. In the first part of the snippet, we create a file object to read the data included in the existing file `my_file.txt`.⁵ At least, we have to pass one argument to `open`: the path pointing to the file. A second optional argument is `mode`, which specifies the mode in which the file is opened to source. It defaults to `r`, which means open for reading in text mode. Other common values are `w` for writing,⁶ `x` for exclusive creation, and `a` for appending.⁷ To read a file’s contents, we use the `.read()` method (see line 9), returning a string object (see line 10).

Snippet 4.24 — data input with open

```

1  # create a pipe to a file
2  >>> file = open(file="my_file.txt", mode="r")
3
4  # calling "file" yields the attributes of the file object
5  >>> file
6  <_io.TextIOWrapper name="my_file.txt" mode="r" encoding="UTF-8">
7
8  # let us source the data
9  >>> data = file.read()
10 >>> print(data)
11 Hi there
12
13 # close the pipe
14 >>> file.close()
```

How do we write the data in the current Python session to a file?

Snippet 4.25 illustrates how to use `open` for data writing. In the first part of the snippet, we create three strings — i.e., the information we are manipulating in the active Python session (see lines 2, 4, and 6). Then, we create a file object in ‘writing’ mode (see the value passed to `mode`, line 9). Finally, we manipulate the three strings (as a sample task, in line 12, we concatenate `FIRSTLAW`, `SECONDLAW`, `THIRDLAW`) and write the result to a file (line 16).

Snippet 4.25 — data output with `open`

```
1 # the strings (data) to save permanently to a file
2 >>> FIRSTLAW = "A robot may not injure a human being or, through inaction, \"
3             \"allow a human being to come to harm.\"
4 >>> SECONDLAW = "A robot must obey the orders given it by human beings except \"
5             \"where such orders would conflict with the First Law.\"
6 >>> THIRDLAW = "A robot must protect its own existence as long as such \"
7             \"protection does not conflict with the First or Second Law.\"
8
9 # create a pipe to a file
10 >>> file = open(file="my_file.txt", mode="w")
11
12 # concatenate the strings
13 >>> TO_WRITE = "\n".join([FIRSTLAW, SECONDLAW, THIRDLAW])
14
15 # write the concatenated strings
16 >>> file.write(TO_WRITE)
17
18 # close the pipe
19 >>> file.close()
```

How about reading a single line from a file?

Hold on: what is a line? A string whose last character is `\n`. We can read a single line from a file using the `.readline()` method (see Snippet 4.26). Such a method starts by reading the first line included in the file (see line 11); then, it reads any subsequent lines included in the file (see line 15); when it reaches the end of the file (EOF), it returns the empty string `""` (see line 19).

Snippet 4.26 — reading one line at a time with `.readline()`

```

1 # the strings (data) to save permanently to a file
2 >>> DATA = "The first line\nThe second line"
3
4 # create a pipe to a file and write DATA
5 >>> file = open(file="my_file.txt", mode="w")
6 >>> file.write(DATA)
7 >>> file.close()
8
9 # read one line from the file
10 >>> file = open(file="my_file.txt", mode="r")
11 >>> file.readline()
12 "The first line\n"
13
14 # calling file.readline() again reads the subsequent line
15 >>> file.readline()
16 "The second line"
17
18 # ... and so on until the end of the file is reached
19 >>> file.readline()
20 ""

```

How about reading multiple lines at a time?

The `.readlines()` method reads the lines from a file and returns them as a list (see Snippet 4.27).

Snippet 4.27 — reading multiple lines at a time with `.readlines()`

```

1 # the strings (data) to save permanently to a file
2 >>> DATA = "A\nB\nC\nD"
3
4 # create a pipe to a file and write DATA
5 >>> file = open(file="my_file.txt", mode="w")
6 >>> file.write(DATA)
7 >>> file.close()
8
9 # read multiple lines
10 >>> file = open(file="my_file.txt", mode="r")
11 >>> file.readlines()
12 ['A\n', 'B\n', 'C\n', 'D']
13
14 # equivalently to the previous line, we can use 'list'
15 ['A\n', 'B\n', 'C\n', 'D']

```

What are the most common file methods?

Table 4.9 illustrates some key file methods' names and their corresponding synopsis.

4.8 Python Statements and Syntax

TABLE 4.9
Popular file methods

Method	Description
<code>file.close()</code>	Closes the file
<code>file.detach()</code>	Returns the separated raw stream from the buffer
<code>file.fileno()</code>	Returns a number that represents the stream as per the OS' perspective
<code>file.flush()</code>	Flushes the internal buffer
<code>file.isatty()</code>	Returns whether the file stream is interactive or not
<code>file.read()</code>	Returns the file content
<code>file.readable()</code>	Returns whether the file stream can be read or not
<code>file.readline()</code>	Returns one line from the file
<code>file.readlines()</code>	Returns a list of lines from the file
<code>file.seek()</code>	Change the file position
<code>file.seekable()</code>	Returns whether the file allows us to change the file position
<code>file.tell()</code>	Returns the current file position
<code>file.truncate()</code>	Resizes the file to a specified size
<code>file.writable()</code>	Returns whether the file can be written to or not
<code>file.write()</code>	Writes the specified string to the file
<code>file.writelines()</code>	Writes a list of strings to the file

Notes: `file` is a fictionary object used to illustrate the usage of the file methods.

What is a Python statement?

In his popular book ‘Learning Python,’ Lutz provides a concise and effective description of what a Python statement is:

In simple terms, statements are the things you write to tell Python what your programs should do. If, as suggested [omitted], programs “do things with stuff,” then statements are the way you specify what sort of things a program does. Less informally, Python is a procedural, statement-based language; by combining statements, you specify a procedure that Python performs to satisfy a program’s goals.

What are Python’s statements?

Table 4.10 illustrates common Python statements, their role, and application examples. Some of these statements were used in the examples considered so far. Other statements — the majority — will be faced in the next sections of the current chapter and/or in the subsequent chapters.

TABLE 4.10
Python Statements

Statement	Role	Example
<code>import</code>	Module access	<code>import math</code>
<code>from</code>	Attribute access	<code>from math import sqrt</code>
<code>class</code>	Building ad hoc objects	<code>class Subclass(Superclass):</code> <code>def method(self):</code> <code>pass</code>
<code>del</code>	Deleting references	<code>del a</code>
Assignment	Creating references	<code>a = "before b"</code>
Calls and other expressions	Running functions	<code>file.write("Hello")</code>
<code>print</code>	Printing objects	<code>print("Hello")</code>
<code>if/elif/else</code>	Selecting actions	<code>if "abc" in text:</code> <code>print(text)</code>
<code>for/else</code>	Iteration	<code>for x in mylist:</code> <code>print(x)</code>
<code>while/else</code>	General loops	<code>while X > Y:</code> <code>print("Hello")</code>
<code>pass</code>	Empty placeholder	<code>while True:</code> <code>pass</code>
<code>break</code>	Loop exit	<code>while True:</code> <code>if exit test():</code> <code>break</code>
<code>continue</code>	Loop continue	<code>while True:</code> <code>if skip test():</code> <code>continue</code>
<code>def</code>	Functions and methods	<code>def f(a, b, c=1, *d):</code> <code>print(a+b+c+d[0])</code>
<code>return</code>	Functions results	<code>def f(a, b, c=1, *d):</code> <code>return a+b+c+d[0]</code>
<code>yield</code>	Generator functions	<code>def gen(n):</code> <code>for i in n:</code> <code>yield i*2</code>

4.9 Control Flow (or If-Then Statements)

What is control flow in Python?

Many Python statements we write are compound statements: there is one statement nested inside another. The outer statement is called the ‘if’ statement and the inner statement is called the ‘then’ statement. The ‘if’ statement is used to determine whether to execute the ‘then’ statement. Specifically, the ‘then’ statement is executed insofar as the ‘if’ statement evaluates to ‘True.’ Snippet 4.28 illustrates a control flow case, a simple rule-based product recommender that suggests products based on users’ purchasing patterns. If product `x` belongs to a user’s set of past purchases, then a certain item is recommended; otherwise, no recommendation is offered (see lines 7 and 8, containing the `else` statement).

Snippet 4.28 — an example of control flow in Python

```
1 # a set with a customer's past purchases
2 >>> S = set(["a", "x", "u"])
3
4 # a rule-based product recommender
5 >>> if "x" in S:
6 ...     print("Customers who bought x also bought Air Jordan 7 Retro Miro")
7 ... else:
8 ...     pass
9 Customers who bought x also bought Air Jordan 7 Retro Miro
```

End of line → end of the statement

Any Python statements are contained in the same line — the end of the line equates to the statement end. In the interest of redundancy, Python statements do not traverse multiple lines. In Snippet 4.28, the if statement is in line 5; the then statement is in line 6.

The ‘:’ character is required for nested statements!

The colon character is required to separate the if statement from the ‘then’ statement (see Snippet 4.28 line 5).

Indentation has substantive meaning in Python!

‘Then’ statements are indented (with a tab or four consecutive spaces). Do not creatively use indents to embellish your code — that is not consistent with Python’s rules and design principles (see Snippet 4.28 line 6).

End of indentation → end of nested statements

Multiple ‘then’ statements

‘Then’ statements are indented (with a tab or four consecutive spaces). In Snippet 4.28, the indentation in line 6 makes lines 5 and 6 to be evaluated together.

Exmple 4.29 shows how to use `elif` to concatenate multiple ‘if-then’ statements in the same control flow. Like in Snippet 4.28, the ‘else’ statement defines the residual behavior of the control flow; that is, what Python does when both the ‘if’ and ‘elif’ statements evaluate to ‘False.’

Snippet 4.29 — an example of control flow with multiple ‘if-then’ statements in Python

```

1 # a list with a customer's past purchases
2 >>> S = set(["a", "w", "u"])
3
4 # a rule-based product recommender
5 >>> if "x" in S:
6 ...     print("Customers who bought x also bought Air Jordan 7 Retro Miro")
7 ... elif "w" in S:
8 ...     print("How about Converse Chuck Taylor All Star?")
9 ... else:
10 ...     print("Falling short of suggestions --- I'm a dull recommender!")
11 How about Converse Chuck Taylor All Star?
```

Nested if-then statements

In Python, it is possible to nest an if-then statement into another. In Snippet 4.30, the ‘if’ statements in lines 6 and 8 are nested inside the ‘if’ statement in line 5. It is worth noticing that lines 7 and 9 — regarding ‘then’ statements — are indented twice because they terminate distinct if-then statements nested in the broader if-then statement commencing on line 5.

Snippet 4.30 — an example of nested control flow in Python

```

1 # a list with a customer's past purchases
2 >>> S = set(["a", "x", "b"])
3
4 # a rule-based product recommender
5 >>> if "x" in S:
6 ...     if "a" in S:
7 ...         print("Customers who bought x & a also bought Air Force")
8 ...     elif "u" in S:
9 ...         print("Customers who bought x & u also bought Air Max 95")
10 ... else:
11 ...     print("Falling short of suggestions --- I'm a dull recommender!")
12 Customers who bought x & a also bought Air Force
```

4.10 While and For Loops

Looping!

How do we write loops in Python?

What is the difference between `for` and `while` statements?

Oftentimes, we write Python statements that repeat a same task — i.e., they loop a certain number of times or over multiple items.

Using `for` and `while` statements.

The `while` statement provides a way to code general loops. The `for` statement is designed for stepping through the items in a sequence or other iterable object and running a block of code for each.

Snippet 4.31 — while loop examples

```
1 # loop until reaching a numeric threshold
2 >>> i = 0
3 >>> while i <= 3:
4 ...     print(i)
5 ...     i = i + 1
6 0
7 1
8 2
9 3
10
11 # loop until an empty string is returned
12 >>> x = "Indiana Jones"
13 >>> while x != "":
14 ...     print(x)
15 ...     x = x[1:]
16 ...
17 Indiana Jones
18 ndiana Jones
19 diana Jones
20 iana Jones
21 ana Jones
22 na Jones
23 a Jones
24 Jones
25 Jones
26 ones
27 nes
28 es
29 s
```

while loops in action

`while` statements run a code block insforar as a test evaluates to `True`. In the upper section of Exmple 4.31, we assing a `i` to a number. Then we create a for loop with the following elements: the first one is a statementing testing wether `i` is smaller or equal 3 (see line 3); the second element is the loop body (indented), which is repeated as long as the test evaluates to `True`. It is worth noticing that every iteration of the loop body produces a unitary increase in `i` — therefore, the program leaves the loop after four iterations. In the lower section of Snippet 4.31, we assigning the variable `x` to a string (line 13), which we eventually print (line 14) and slice (line 15) until we get an empty string (line 13x).

for loops in action

`for` steps through a sequence of items and carries out a task. In the upper section of Snippet 4.32, we print the result of a mathematical operation that is deployed over the items of a `list` (an example of Python iterable object). The code included in line 2 assigns temporarily the variable `item` to an element of iterable. Then, the code block (indented) is executed over the temporary object. In the lower section of Snippet 4.32, the execution of the loop operates a mathematical expression over the items of a first list and appends the outcome to a second list (line 10).

Snippet 4.32 — a for loop example

```

1  # print the result of a mathematical operation carried out over a list of items
2  >>> for item in [0, -99, 13, 6.54]:
3  ...     print(item ** 0.5)
4
5  # run a mathematical operation on a list of items and append the outcome
6  # to a second list
7  >>> input = [2, 8, 1]
8  >>> output = []
9  >>> for item in input:
10 ...     output.append(item + 1)
11 >>> print(output)
12 [3, 9, 2]
```

for loops with dictionaries

Like lists, dictionaries are iterable objects. In the upper section of Snippet 4.33, we create a dictionary and iterate over its items printing a simple predicate. As we know from section 4.4, we access a dictionary's values by keys. Hence, in line 5, we retrieve the keys of `D`. Then, in line 12, we fetch the value of the temporary object `k`, namely, `D[k]`. Particularly, we print the temporary object `k`, the string object `IS`, and the value associated with `k`; that is, `D[k]`. In Snippet 34, we accomplish the same task of Snippet 33. However, the loop regards a dictionary's items — i.e., key-value pairs — instead of keys (that is self-evident from the comparison of Snippet 33's line 55 and Snippet 34's line 5).

Snippet 4.33 — looping on dictionary keys

```
1 # the dictionary
2 >>> D = {"Thor": "Asgardian", "Vision": "android", "Wanda Maximoff": "human"}
3
4 # get the keys of D
5 >>> keys = D.keys()
6 >>> print(keys)
7 dict_keys(['Thor', 'Vision', 'Wanda Maximoff'])
8
9 # iterate over the keys to fetch the dictionary values and do something
10 # with them
11 >>> for k in keys:
12 ...     print(k + " IS " + D[k])
13 Thor IS Asgardian
14 Vision IS android
15 Wanda Maximoff IS human
```

Snippet 4.34 — looping on dictionary items

```
1 # the dictionary
2 >>> D = {"Thor": "Asgardian", "Vision": "android", "Wanda Maximoff": "human"}
3
4 # get the items of D
5 >>> items = D.items()
6
7 # iterate over key-value pairs and do something with them
8 >>> for k, v in items:
9 ...     print(k + " IS " + v)
10 Thor IS Asgardian
11 Vision IS android
12 Wanda Maximoff IS human
```

Counter for loops

The built-in class `range` provides an immutable sequence that is particularly helpful for loops that repeat an action a certain number of times. Snippet 35 shows an example of a for loop with `range`.

Snippet 4.35 — a counter loop example

```
1 # show the outcome of range
2 >>> list(range(3))
3 [0, 1, 2]
4
5 # use range in a for loop
6 >>> for i in range(3):
7 ...     print(i, ":-)")
8 0 :-)
9 1 :-)
10 2 :-)
```

Nested for loops

A Python statement that contains multiple for loops is a nested for loop. Mainly, a for loop allows to carry out a task over the elements of two iterables jointly. The outer loop considers the individual items of the first iterable (see Snippet 4.26, line 6); the inner loop (indented) considers the individual items of the second list (see line 7). Once we have created a pair of temporary objects, we can do something with it (see line 8).

Snippet 4.36 — a nested loop example

```
1 # the lists
2 >>> LETTERS = ["x", "y", "z"]
3 >>> COLORS = ["blue", "green", "red"]
4
5 # create all permutations of letters and colors and print them
6 >>> for i in LETTERS:
7 ...     for j in COLORS:
8 ...         print(i, " - ", j)
9 x <-> blue
10 x <-> green
11 x <-> red
12 y <-> blue
13 y <-> green
14 y <-> red
15 z <-> blue
16 z <-> green
17 z <-> red
```

Nested for loops Vs. `zip` for loops

Contrarily to the nested `for` loops, which considers all permutations containing multiple iterables' items, the built-in `zip` steps through several iterables *in parallel*, producing tuples with an item from each one. As shown in Snippet 4.37, there is neither an inner nor an outer `for` loop in this case — instead, there is a single loop considering two temporary objects, `i` and `j`, that occupy the same position in the offset of the iterables at hand (see line 6; the first item from the first iterable goes with the first item from the second iterable, the second item from the first iterable goes with the second item from the second iterable, and so on and so forth).

Snippet 4.37 — looping over two iterables in parallel with `zip`

```
1 # the lists
2 >>> LETTERS = ["x", "y", "z"]
3 >>> COLORS = ["blue", "green", "red"]
4
5 # create one-to-one matches of items and do something with them
6 >>> for i, j in zip(LETTERS, COLORS):
7     ...     print(i, " <-> ", j)
8 x <-> blue
9 y <-> green
10 z <-> red
```

4.11 Iterations and Comprehensions

Over and beyond `while` and `for` loops

As we know from the previous section, `while` and `for` loops can handle most repetitive tasks programs need to perform. However, Python provides further tools to make loops *easier to write/read* and *more efficient*. One of the most prominent tools is `list comprehension`.

Why do we use `list comprehensions`?

To create a list containing the outcome of an action repeated over an iterable's items (see Snippet 4.38, line 14).

How do we create `list comprehensions`?

We include a Python statement containing a `for` clause among brackets (see Snippet 4.38, line 12).

Snippet 4.38 — for loop Vs. list comprehension

```

1 # the for loop way
2 # --+ create an empty list
3 L = []
4 # --+ create a for loop appending the square of some items
5 >>> for i in range(3):
6 ...     L.append(i ** 2)
7 # --+ print the list
8 >>> print(L)
9 [0, 1, 4]
10
11 # the list comprehension way
12 >>> L = [i ** 2 for i in range(3)]
13 >>> print(L)
14 [0, 1, 4]

```

How do we implement a nested for loop in list comprehensions?

As Snippet 4.39 shows, a nested for loop becomes a one-liner in a list comprehension. The first for clause in line 7 would correspond to the outer for loop reported in Snippet 4.36, whereas the second for clause in line 7 would correspond to the inner for loop reported in Snippet 4.36.

Snippet 4.39 — nested for loop with list comprehensions

```

1 # the lists
2 >>> LETTERS = ["x", "y", "z"]
3 >>> COLORS = ["blue", "green", "red"]
4
5 # implementing a nested for loop with a list comprehension
6 >>> LETTER2COLOR = ["{} <-> {}".format(i, j) for i in LETTERS for j in COLORS]
7 ['x <-> blue',
8  'x <-> green',
9  'x <-> red',
10 'y <-> blue',
11 'y <-> green',
12 'y <-> red',
13 'z <-> blue',
14 'z <-> green',
15 'z <-> red']

```

Can we use the zip generator within a list comprehension?

Yes, we can. To do that, the for clause must consider two iterables at once (see Snippet 4.40, line 6).

Snippet 4.40 — looping over two iterables in parallel with `zip` and list comprehension

```

1 # the lists
2 >>> LETTERS = ["x", "y", "z"]
3 >>> COLORS = ["blue", "green", "red"]
4
5 # implementing a nested for loop with a list comprehension
6 >>> LETTER2COLOR = ["{} <-> {}".format(i, j) for i, j in zip(LETTERS, COLORS)]
7 ['1 <-> blue', '2 <-> green', '3 <-> red']

```

Can we use embed control flow in a list comprehension?

Yes, we can. To do that, the `for` clause must be preceded by an `if` statement and, at least, an `else` statement (see for example Snippet 4.40's line 22).

Snippet 4.41 — control flow in list comprehensions

```

1 # import the function log from math
2 from math import log
3
4 # the object to manipulate
5 >>> L1 = [0, 1, 2]
6
7 # the for loop way
8 # --+ the empty list
9 L2 = []
10 # --+ the for loop appending the log of some items
11 >>> for i in L1:
12 ...     if i > 0:
13 ...         L2.append(log(i))
14 ...     else:
15 ...         L2.append(log(i + 0.001))
16 # --+ print the list
17 >>> print(L2)
18 [-6.907755278982137, 0.0, 0.6931471805599453]
19
20 # the list comprehension way
21 # --+ the list comprehension is a one-liner!
22 >>> L2 = [log(i) if i > 0 else log(i + 0.001) for i in L1]
23 # --+ print the list
24 >>> print(L2)
25 [-6.907755278982137, 0.0, 0.6931471805599453]

```

Notes

¹Lutz, Mark. *Learning Python: Powerful object-oriented programming*. O'Reilly Media, Inc., 2013.

²Floating numbers are stored in binaries with an assigned level of precision that is typically equivalent to 15 or 16 decimals.

³As per the documentation of the Python programming language, `math` cannot be used with complex numbers.

⁴The official Python documentation has an extensive section on operator precedence rules in the section dedicated to [syntax of expressions](#)

⁵For the sake of simplicity, we assume the target file is located in the same directory as the Python script.

⁶By default, `w` truncates the file if it already exists

⁷If encoding is not specified the encoding used is platform-dependent. Specifically, `locale.getpreferredencoding(False)`

is called to get the current locale encoding. Character encoding is the process of assigning numbers to graphical characters, especially the written characters of human language, allowing them to be stored, transmitted, and transformed using digital computers.

Chapter 5

Technical & Scientific Computation with NumPy and SciPy

Learning goals:

- -
-

5.1 Installing NumPy and SciPy

Do NumPy and SciPy come with the official Python installation file?

No, they do not. You need to install them separately using the package manager `pip`.

I am an Anaconda user: do I need to install NumPy and SciPy separately?

No, you do not. NumPy and SciPy are included in ‘base’, the default environment of Anaconda. However, if you create a new environment, you need to install NumPy and SciPy — as well as all the other modules you need — with the package manager `conda`.

How do I install NumPy and SciPy?

The easiest way is using the command line. Anaconda users run `$ conda install numpy scipy`, whereas Python official release users run `$ pip install numpy scipy`. Having said that, Anaconda users can also install the modules they need from within Anaconda-Navigator.

5.2 NumPy ndarray

What is a NumPy ndarray?

Put simply, an `ndarray` is a data container, like dictionaries and lists.

Can `ndarrays` contain objects of different type?

How do we create an `ndarray`?

No, they cannot. An `ndarray` must contain homogeneous items; that is, items of the same type.

As shown in Snippet 5.1, we pass an object to `numpy.array`. If the object we pass is a scalar, a 0-dimensional array containing object is returned (line 5). Passing a list to `numpy.array` produces a one-dimensional array (line 9); passing a list of lists produces a two-dimensional array (line 13); finally, passing a list of lists of lists produces a three-dimensional array (line 18).

Snippet 5.1 — creating an `ndarray`

```

1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
3
4 # a 0-D array
5 >>> np.array(0)
6 array(0)
7
8 # a 1-D array
9 >>> np.array([1, 2, 3, 4])
10 array([1, 2, 3, 4])
11
12 # a 2-D array
13 >>> np.array([[1, 2], [3, 4]])
14 array([[1, 2],
15        [3, 4]])
16
17 # a 3-D array
18 >>> np.array([[[1, 2], [3, 4]], [[5, 6], [7, 8]]])
19 array([[[1, 2],
20        [3, 4]],
21       [[5, 6],
22        [7, 8]]])
23
```

What are the distinctive features of `ndarrays`?

`ndarrays` have been designed — and tuned over time — with flexibility and efficiency in mind. For example, `ndarrays` allow to carry out computations on arrays with a syntax similar to scalar values. As shown in Snippet 5.2, we can multiply an array by a scalar (line 10) as well as add two vectors (14). That is not possible if we use pure Python code. Multiplying a list by a scalar N replicates the ordered collection of items N times (see line 23). Adding two lists yields concatenation (see line 30). At the same time, `ndarrays` support the analysis of large volumes of data⁸.

Snippet 5.2 — NumPy allows to manipulate vectors with an expressive syntax

```

1 # import numpy with the socially accepted alias 'np'
2 import numpy as np
3
4 # generate some random
5 >>> DATA = np.random.randn(3)
6 >>> print(DATA)
7 [-0.44144029 -0.44451097  0.31997294]
8
9 # can we multiply a list by a scalar with NumPy? Of course!
10 >>> print(DATA * 3)
11 [-4.4144029 , -4.44510974,  3.19972941]
12
13 # can we sum two arrays with NumPy? Of course!
14 >>> print(DATA + DATA)
15 [-0.88288058, -0.88902195,  0.63994588]
16
17 # let us try to replicate the previous tasks in pure Python?
18 # --+ get the DATA as a list
19 >>> DATA = list(DATA)
20 >>> print(DATA)
21 [-0.4414402896845323, -0.4445109735283278, 0.31997294069261617]
22 # --+ is the NumPy syntax of line 10 still valid if we use a list? Nope
23 >>> print(DATA * 3)
24 [
25 -0.4414402896845323, -0.4445109735283278, 0.31997294069261617,
26 -0.4414402896845323, -0.4445109735283278, 0.31997294069261617,
27 -0.4414402896845323, -0.4445109735283278, 0.31997294069261617
28 ]
29 # --+ is the NumPy syntax of line 14 still valid if we use a list? Nope
30 >>> print(DATA + DATA)
31 [
32 -0.4414402896845323, -0.4445109735283278, 0.31997294069261617,
33 -0.4414402896845323, -0.4445109735283278, 0.31997294069261617
34 ]

```

How do I check an array's number of dimensions?

In Snippet 5.3, we saw NumPy infers an `ndarrays`'s number of dimensions from the data. Snippet 5.3 shows how to access `.ndim`, the `ndarrays`'s attribute concerning the number of dimensions. In the example, `DATA` has two dimensions (e.g., coordinates).

How do I check an array's shape?

The lower section of Snippet 5.3 shows how to access `.shape`, the `ndarrays`'s attribute concerning the shape. In the example, each dimension of the `DATA` has size 2.

Snippet 5.3 — an array's number of dimensions and shape

```

1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
3
4 # the data
5 >>> DATA = np.array([[1, 2], [3, 4]])
6 >>> print(DATA)
7 [[1 2]
8  [3 4]]
9
10 # get the number of dimensions
11 >>> DATA.ndim
12 2
13
14 # get the shape
15 >>> DATA.shape
16 (2, 2)

```

What are the attributes of an array?

Table 5.1 illustrates the common use attributes of `ndarrays`.

TABLE 5.1
Common Use Attributes of NumPy `array`

Attribute	Synopsis
<code>DATA.flags</code>	Information about the memory layout of the array
<code>DATA.shape</code>	Tuple of array dimensions
<code>DATA.strides</code>	Tuple of bytes to step in each dimension when traversing an array
<code>DATA.ndim</code>	Number of array dimensions
<code>DATA.data</code>	Python buffer object pointing to the start of the array's data
<code>DATA.size</code>	Number of elements in the array
<code>DATA.itemsize</code>	Length of one array element in bytes
<code>DATA.nbytes</code>	Total bytes consumed by the elements of the array
<code>DATA.dtype</code>	Data-type of the array's elements

Notes: `DATA` is a fictional object used to illustrate the usage of the `array` attributes.

I know that a NumPy array must contain homogenous data — but which object types are allowed?

How do I specify the data type of an array?

Table 5.2 reports the NumPy data types. Python beginners are not supposed to appreciate the distinctive attributes of each type. Instead, they may want to get a clear understanding of the high-level types, namely, floating points, complex, integer, boolean, string, or general Python objects. When it comes to working on sophisticated projects, requiring more control over the storage types, then it is highly suggested to get a thorough knowledge of the types in Table 5.2. It is worth noticing that `dtypes` are a source of NumPy’s flexibility for interacting with data coming from other systems. In most cases, they provide a mapping directly onto an underlying disk or memory representation, which makes it easy to read and write binary streams of data to disk and also to connect to code written in a low-level language like C or Fortran. The numerical `dtypes` are named the same way: a type name, like `float` or `int`, followed by a number indicating the number of bits per element. A standard double-precision floating-point value takes up 8 bytes or 64 bits. Thus, this type is known in NumPy as `float64`.

As Snippet 5.4 shows, `dtype` is an optional argument of `ndarray`. It is also possible to change `dtype` using the method `.astype()`.

Snippet 5.4 — specifying a nd changing dtype

```

1  # import numpy with the socially accepted alias 'np'
2  >>> import numpy as np
3
4  # accept the default type
5  >>> A = np.array([1, 2, 3, 4, 5])
6
7  # check the type
8  >>> A.dtype
9  dtype('int64')
10
11 # specify the type
12 >>> A = np.array([1, 2, 3, 4, 5], dtype=np.int32)
13
14 # check the type
15 >>> A.dtype
16 dtype('int32')
17
18 # type change
19 >>> S = np.array(['1.25', '-9.6', '42'], dtype=np.string_)
20 >>> S = S.astype(float)
21 >>> S.dtype
22 dtype('float64')
```

TABLE 5.2
NumPy Data Types

Type	Type Code	Synopsis
int8, uint8	i1, u1	Signed and unsigned 8-bit (1 byte) integer types
int16, uint16	i2, u2	Signed and unsigned 16-bit integer types
int32, uint32	i4, u4	Signed and unsigned 32-bit integer types
int64, uint64	i8, u8	Signed and unsigned 64-bit integer types
float16	f2	Half-precision floating point float32 f4 or f Standard single-precision floating point; compatible with C float
float64	f8 or d	Standard double-precision floating point; compatible with C double and Python
float object	float128 f16 or g	Extended-precision floating point
complex64, complex128, complex256	c8, c16, c32	Complex numbers represented by two 32, 64, or 128 floats, respectively
bool	?	Boolean type storing True and False values
object	O	Python object type; a value can be any Python object
string	S	Fixed-length ASCII string type (1 byte per character); for example, to create a string dtype with length 10, use 'S10'
unicode	U	Fixed-length Unicode type (number of bytes platform specific); same specification semantics as string- (e.g., 'U10')

5.3 Array Creation and Manipulation Methods

Does NumPy offer recipes for creating arrays?

Yes, it does. NumPy has seven families of array-creating routines:

- From shape or value
- From existing data
- Creating record arrays (`np.rec`)
- Creating character arrays (`np.char`)
- Numerical ranges
- Building matrices
- The Matrix class

Creating arrays from shape or value

This family of routines creates arrays with a certain number of dimensions, shape, values, and attributes. Snippet 5.5. shows how to create:

- an array with a certain shape and a constant scalar (see lines 5 — `.zeros`, 12 — `.ones`, 19 — `full`), and
- an array containing an identity matrix (see lines 26 — `.eye` — and 33 — `.identity`)

Snippet 5.5 — creating arrays from shape or value

```

1  # import numpy with the socially accepted alias 'np'
2  >>> import numpy as np
3
4  # create an array with zeros only
5  >>> np.zeros([4,4])
6  array([[0., 0., 0., 0.],
7         [0., 0., 0., 0.],
8         [0., 0., 0., 0.],
9         [0., 0., 0., 0.]])
10
11 # create an array with ones only
12 >>> np.ones((4,4))
13 array([[1., 1., 1., 1.],
14        [1., 1., 1., 1.],
15        [1., 1., 1., 1.],
16        [1., 1., 1., 1.]])
17
18 # create a full matrix with a given scalar
19 >>> np.full((4,4), -99)
20 array([[-99, -99, -99, -99],
21        [-99, -99, -99, -99],
22        [-99, -99, -99, -99],
23        [-99, -99, -99, -99]])
24
25 # create an identity array of a given shape with .eye
26 >>> np.eye(4, 3)
27 array([[1., 0., 0.],
28        [0., 1., 0.],
29        [0., 0., 1.],
30        [0., 0., 0.]])
31
32 # create an identity array with .identity
33 >>> np.identity(4)
34 array([[1., 0., 0., 0.],
35        [0., 1., 0., 0.],
36        [0., 0., 1., 0.],
37        [0., 0., 0., 1.]])

```

TABLE 5.3
Routines for Creating Arrays from Shape or Value

Routine	Synopsis
<code>np.empty(shape[, dtype, order, like])</code>	Return a new array of given shape and type, without initializing entries
<code>np.empty_like(prototype[, dtype, order, subok, ...])</code>	Return a new array with the same shape and type as a given array
<code>np.eye(N[, M, k, dtype, order, like])</code>	Return a 2-D array with ones on the diagonal and zeros elsewhere
<code>np.identity(n[, dtype, like])</code>	Return the identity array
<code>np.ones(shape[, dtype, order, like])</code>	Return a new array of given shape and type, filled with ones
<code>np.ones_like(a[, dtype, order, subok, ...])</code>	Return an array of ones with the same shape and type as a given array
<code>np.zeros(shape[, dtype, order, like])</code>	Return a new array of given shape and type, filled with zeros
<code>np.full(shape, fill_value[, dtype, order, like])</code>	Return a new array of given shape and type, filled with fill value
<code>full_like(a, fill_value[, dtype, order, ...])</code>	Return a full array with the same shape and type as a given array

Creating arrays from existing data

In Snippets 5.1 – 5.4, we saw how to use `array` for passing data to a NumPy array. Further routines to create NumPy arrays from existing data include:

- `.fromfunction`, creating an array by executing a function over each coordinate
- `.fromfile`, creating an array from data in a text or binary file
- `.loadtxt`, loading data from a text file

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```

TABLE 5.4
Routines for Creating Arrays from Existing Data

Routine	Synopsis
<code>np.array(object[,dtype, copy, subok, ...])</code>	Create an array
<code>np.asarray(a[, dtype, order, like])</code>	Convert the input to an array
<code>np.asanyarray(a[, dtype, order, like])</code>	Convert the input to an ndarray, but pass ndarray subclasses through
<code>np.ascontiguousarray(a[, dtype, like])</code>	Return a contiguous array (ndim ≥ 1) in memory (C order)
<code>np.asmatrix(data[, dtype])</code>	Interpret the input as a matrix
<code>np.copy(a[, order, subok])</code>	Return an array copy of the given object
<code>np.frombuffer(buffer[, dtype, count, offset, like])</code>	Interpret a buffer as a 1-dimensional array
<code>np.fromfile(file[, dtype, count, sep, offset, like])</code>	Construct an array from data in a text or binary file
<code>np.fromfunction(function, shape, *[dtype, like])</code>	Construct an array by executing a function over each coordinate
<code>np.fromiter(iter, dtype[, count, like])</code>	Create a new 1-dimensional array from an iterable object
<code>np.fromstring(string[, dtype, count, like])</code>	A new 1-D array initialized from text data in a string
<code>np.loadtxt(fname[, dtype, comments, delimiter, ...])</code>	Load data from a text file

Creating arrays from existing data

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```

TABLE 5.5
Routines for Creating Record Arrays

Routine	Synopsis

Creating arrays from existing data

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```


TABLE 5.6
Routines for Creating Character Arrays

Routine	Synopsis

Creating arrays from existing data

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```

TABLE 5.7
Routines for Numerical Ranges

Routine	Synopsis

Creating arrays from existing data

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```

TABLE 5.8
Routines for Building Matrices

Routine	Synopsis

Creating arrays from existing data

Snippet 5.6 — creating arrays from existing data

```
1 # import numpy with the socially accepted alias 'np'
2 >>> import numpy as np
```

TABLE 5.9
Routines for the Matrix Class

Routine	Synopsis

5.4 Universal Functions in NumPy

What is a universal function?

A universal function, or **ufunc**, is a function that performs element-wise operations on **ndarrays**. You can think of them as fast vectorized wrappers for simple functions that take one or more scalar values and produce one or more scalar results⁹.

What is the rationale for using **ufuncs**?

Per the previous point, using a **ufunc** offer substantial performance advantages *vis a' vis* non-vectorized code — i.e., code using built-in Python iterators.

What are the **ufunc** options available in NumPy?

There are circa sixty universal functions implemented in NumPy. For the sake of convenience, the full list of **ufunc** options is reported in Tables A.3 - A.6. The following sections of the current chapter will illustrate how to use some of the popular central **ufuncs** in NumPy.

5.5 Mathematical Methods

5.6 Statistical Methods

5.7 Linear Algebra

5.8 Pseudorandom Number Generation

5.9 File Input and Output with **ndarrays**

5.10

Notes

⁸The Internet has many blog posts showing the performance of NumPy in linear algebra tasks is comparable to compiled languages, such as C.

⁹Here is an interesting passage from the official **NumPy** documentation: *NumPy hands off array processing to C, where looping and computation are much faster than in Python. To exploit this, programmers using NumPy eliminate Python loops in favor of array-to-array operations. vectorization can refer both to the C offloading and to structuring NumPy code to leverage it.*

Chapter 6

Data Management with Pandas

...

Chapter 7

Coda

...

Appendices

Appendix A

Cheat Sheets

TABLE A.1
Helpful Escapes

Escape	Meaning
\\	Backslash (stores one \)
\'	Single quotes escape (stores ')
\"	Double quotes escape (stores ")
\a	Bell
\b	Backspace
\f	Formfeed
\n	Newline
\r	Carriage return
\t	Horizontal tab
\v	Vertical tab

TABLE A.2
Comprehensive List of String Methods

Cases I	
s.capitalize()	Capitalize s # 'hello' => 'Hello'
s.lower()	Lowercase s # 'HELLO' => 'hello'
s.swapcase()	Swap cases of all characters in s # 'Hello' => "hELLO"
s.title()	Titlecase s # 'hello world' => 'Hello World'
s.upper()	Uppercase s # 'hello' => 'HELLO'
Sequence Operations I	
s2 in s	Return true if s contains s2
s + s2	Concat s and s2
len(s)	Length of s
min(s)	Smallest character of s
max(s)	Largest character of s
Sequence Operations II	
s2 not in s	Return true if s does not contain s2
s * integer	Return integer copies of s concatenated # 'hello' => 'hellohellohello'
s[index]	Character at index of s
s[i:j:k]	Slice of s from i to j with step k
s.count(s2)	Count of s2 in s
Whitespace I	
s.center(width)	Center s with blank padding of width # 'hi' => ' hi '
s.isspace()	Return true if s only contains whitespace characters
s.ljust(width)	Left justify s with total size of width # 'hello' => 'hello '
s.rjust(width)	Right justify s with total size of width # 'hello' => ' hello'
s.strip()	Remove leading and trailing whitespace from s # ' hello ' => 'hello'
Find / Replace I	
s.index(s2, i, j)	Index of first occurrence of s2 in s after index i and before index j
s.find(s2)	Find and return lowest index of s2 in s
s.index(s2)	Return lowest index of s2 in s (but raise ValueError if not found)
s.replace(s2, s3)	Replace s2 with s3 in s
s.replace(s2, s3, count)	Replace s2 with s3 in s at most count times
s.rfind(s2)	Return highest index of s2 in s
s.rindex(s2)	Return highest index of s2 in s (raise ValueError if not found)
Cases II	
s.casefold()	Casefold s (aggressive lowercasing for caseless matching) # 'ßorat' => 'ssorat'
s.islower()	Return true if s is lowercase
s.istitle()	Return true if s is titlecased # 'Hello World' => true
s.isupper()	Return true if s is uppercase
Inspection I	
s.endswith(s2)	Return true if s ends with s2
s.isalnum()	Return true if s is alphanumeric
s.isalpha()	Return true if s is alphabetic
s.isdecimal()	Return true if s is decimal
s.isnumeric()	Return true if s is numeric
s.startswith(s2)	Return true if s starts with s2

TABLE A.2
(Cont'ed)

Splitting I	
s.join('123')	Return s joined by iterable '123' # 'hello' =>'1hello2hello3'
s.partition(sep)	Partition string at sep and return 3-tuple with part before, the sep itself, and part after # 'hello' =>('he', 'l', 'lo')
s.rpartition(sep)	Partition string at last occurrence of sep, return 3-tuple with part before, the sep, and part after # 'hello' =>('hel', 'l', 'o')
s.rsplit(sep, maxsplit)	Return list of s split by sep with rightmost maxsplits performed
s.split(sep, maxsplit)	Return list of s split by sep with leftmost maxsplits performed
s.splitlines()	Return a list of lines in s # 'hello\nworld' =>['hello', 'world']
Inspection II	
s[i:j]	Slice of s from i to j
s.endswith((s1, s2, s3))	Return true if s ends with any of string tuple s1, s2, and s3
s.isdigit()	Return true if s is digit
s.isidentifier()	Return true if s is a valid identifier
s.isprintable()	Return true is s is printable
Whitespace II	
s.center(width, pad)	Center s with padding pad of width # 'hi' =>'padpadhipadpad'
s.expandtabs(integer)	Replace all tabs with spaces of tabsize integer # 'hello\tworld' =>'hello world'
s.lstrip()	Remove leading whitespace from s # ' hello ' =>'hello '
s.rstrip()	Remove trailing whitespace from s # ' hello ' =>'hello'
s.zfill(width)	Left fill s with ASCII '0' digits with total length width # '42' =>'00042'

TABLE A.3
NumPy Universal Functions: Mathematical Operations

Universal Function	Synopsis
<code>add(x1, x2, /[, out, where, casting, order, ...])</code>	Add arguments element-wise
<code>subtract(x1, x2, /[, out, where, casting, ...])</code>	Subtract arguments, element-wise
<code>multiply(x1, x2, /[, out, where, casting, ...])</code>	Multiply arguments element-wise
<code>matmul(x1, x2, /[, out, casting, order, ...])</code>	Matrix product of two arrays
<code>divide(x1, x2, /[, out, where, casting, ...])</code>	Divide arguments element-wise
<code>logaddexp(x1, x2, /[, out, where, casting, ...])</code>	Logarithm of the sum of exponentiations of the inputs
<code>logaddexp2(x1, x2, /[, out, where, casting, ...])</code>	Logarithm of the sum of exponentiations of the inputs in base-2
<code>true_divide(x1, x2, /[, out, where, ...])</code>	Divide arguments element-wise
<code>floor_divide(x1, x2, /[, out, where, ...])</code>	Return the largest integer smaller or equal to the division of the inputs
<code>negative(x, /[, out, where, casting, order, ...])</code>	Numerical negative, element-wise
<code>positive(x, /[, out, where, casting, order, ...])</code>	Numerical positive, element-wise
<code>power(x1, x2, /[, out, where, casting, ...])</code>	First array elements raised to powers from second array, element-wise
<code>float_power(x1, x2, /[, out, where, ...])</code>	First array elements raised to powers from second array, element-wise
<code>remainder(x1, x2, /[, out, where, casting, ...])</code>	Returns the element-wise remainder of division
<code>mod(x1, x2, /[, out, where, casting, order, ...])</code>	Returns the element-wise remainder of division
<code>fmod(x1, x2, /[, out, where, casting, ...])</code>	Returns the element-wise remainder of division
<code>divmod(x1, x2[, out1, out2], / [[, out, ...])</code>	Return element-wise quotient and remainder simultaneously
<code>absolute(x, /[, out, where, casting, order, ...])</code>	Calculate the absolute value element-wise
<code>fabs(x, /[, out, where, casting, order, ...])</code>	Compute the absolute values element-wise
<code>rint(x, /[, out, where, casting, order, ...])</code>	Round elements of the array to the nearest integer
<code>sign(x, /[, out, where, casting, order, ...])</code>	Returns an element-wise indication of the sign of a number
<code>heaviside(x1, x2, /[, out, where, casting, ...])</code>	Compute the Heaviside step function
<code>conj(x, /[, out, where, casting, order, ...])</code>	Return the complex conjugate, element-wise
<code>conjugate(x, /[, out, where, casting, ...])</code>	Return the complex conjugate, element-wise
<code>exp(x, /[, out, where, casting, order, ...])</code>	Calculate the exponential of all elements in the input array
<code>exp2(x, /[, out, where, casting, order, ...])</code>	Calculate 2^{**p} for all p in the input array
<code>log(x, /[, out, where, casting, order, ...])</code>	Natural logarithm, element-wise
<code>log2(x, /[, out, where, casting, order, ...])</code>	Base-2 logarithm of x
<code>log10(x, /[, out, where, casting, order, ...])</code>	Return the base 10 logarithm of the input array, element-wise
<code>expm1(x, /[, out, where, casting, order, ...])</code>	Calculate $\exp(x) - 1$ for all elements in the array
<code>log1p(x, /[, out, where, casting, order, ...])</code>	Return the natural logarithm of one plus the input array, element-wise
<code>sqrt(x, /[, out, where, casting, order, ...])</code>	Return the non-negative square-root of an array, element-wise
<code>square(x, /[, out, where, casting, order, ...])</code>	Return the element-wise square of the input
<code>cbrt(x, /[, out, where, casting, order, ...])</code>	Return the cube-root of an array, element-wise
<code>reciprocal(x, /[, out, where, casting, ...])</code>	Return the reciprocal of the argument, element-wise
<code>gcd(x1, x2, /[, out, where, casting, order, ...])</code>	Returns the greatest common divisor of $\ x1\ $ and $\ x2\ $
<code>lcm(x1, x2, /[, out, where, casting, order, ...])</code>	Returns the lowest common multiple of $\ x1\ $ and $\ x2\ $

TABLE A.4
NumPy Universal Functions: Trigonometric Operations

Universal Function	Synopsis
<code>sin(x, /[, out, where, casting, order, ...])</code>	Trigonometric sine, element-wise
<code>cos(x, /[, out, where, casting, order, ...])</code>	Cosine element-wise
<code>tan(x, /[, out, where, casting, order, ...])</code>	Compute tangent element-wise
<code>arcsin(x, /[, out, where, casting, order, ...])</code>	Inverse sine, element-wise
<code>arccos(x, /[, out, where, casting, order, ...])</code>	Trigonometric inverse cosine, element-wise
<code>arctan(x, /[, out, where, casting, order, ...])</code>	Trigonometric inverse tangent, element-wise
<code>arctan2(x1, x2, /[, out, where, casting, ...])</code>	Element-wise arc tangent of x1/x2 choosing the quadrant correctly
<code>hypot(x1, x2, /[, out, where, casting, ...])</code>	Given the "legs" of a right triangle, return its hypotenuse
<code>sinh(x, /[, out, where, casting, order, ...])</code>	Hyperbolic sine, element-wise
<code>cosh(x, /[, out, where, casting, order, ...])</code>	Hyperbolic cosine, element-wise
<code>tanh(x, /[, out, where, casting, order, ...])</code>	Compute hyperbolic tangent element-wise
<code>arcsinh(x, /[, out, where, casting, order, ...])</code>	Inverse hyperbolic sine element-wise
<code>arccosh(x, /[, out, where, casting, order, ...])</code>	Inverse hyperbolic cosine, element-wise
<code>arctanh(x, /[, out, where, casting, order, ...])</code>	Inverse hyperbolic tangent element-wise
<code>degrees(x, /[, out, where, casting, order, ...])</code>	Convert angles from radians to degrees
<code>radians(x, /[, out, where, casting, order, ...])</code>	Convert angles from degrees to radians
<code>deg2rad(x, /[, out, where, casting, order, ...])</code>	Convert angles from degrees to radians
<code>rad2deg(x, /[, out, where, casting, order, ...])</code>	Convert angles from radians to degrees

TABLE A.5
NumPy Universal Functions: Floating Operations

Universal Function	Synopsis
<code>isfinite(x, /[, out, where, casting, order, ...])</code>	Test element-wise for finiteness (not infinity and not Not a Number)
<code>isinf(x, /[, out, where, casting, order, ...])</code>	Test element-wise for positive or negative infinity
<code>isnan(x, /[, out, where, casting, order, ...])</code>	Test element-wise for NaN and return result as a boolean array
<code>isnat(x, /[, out, where, casting, order, ...])</code>	Test element-wise for NaT (not a time) and return result as a boolean array
<code>fabs(x, /[, out, where, casting, order, ...])</code>	Compute the absolute values element-wise
<code>signbit(x, /[, out, where, casting, order, ...])</code>	Returns element-wise True where signbit is set (less than zero)
<code>copysign(x1, x2, /[, out, where, casting, ...])</code>	Change the sign of x1 to that of x2, element-wise
<code>nextafter(x1, x2, /[, out, where, casting, ...])</code>	Return the next floating-point value after x1 towards x2, element-wise
<code>spacing(x, /[, out, where, casting, order, ...])</code>	Return the distance between x and the nearest adjacent number
<code>modf(x[, out1, out2], / [[, out, where, ...])</code>	Return the fractional and integral parts of an array, element-wise
<code>ldexp(x1, x2, /[, out, where, casting, ...])</code>	Returns $x1 * 2^{x2}$, element-wise
<code>frexp(x[, out1, out2], / [[, out, where, ...])</code>	Decompose the elements of x into mantissa and two's exponent
<code>fmod(x1, x2, /[, out, where, casting, ...])</code>	Returns the element-wise remainder of division
<code>floor(x, /[, out, where, casting, order, ...])</code>	Return the floor of the input, element-wise
<code>ceil(x, /[, out, where, casting, order, ...])</code>	Return the ceiling of the input, element-wise
<code>trunc(x, /[, out, where, casting, order, ...])</code>	Return the truncated value of the input, element-wise

TABLE A.6
NumPy Universal Functions: Comparison Operations

Universal Function	Synopsis
<code>greater(x1, x2, /[, out, where, casting, ...])</code>	Return the truth value of (x1 > x2) element-wise
<code>greater_equal(x1, x2, /[, out, where, ...])</code>	Return the truth value of (x1 >= x2) element-wise
<code>less(x1, x2, /[, out, where, casting, ...])</code>	Return the truth value of (x1 < x2) element-wise
<code>less_equal(x1, x2, /[, out, where, casting, ...])</code>	Return the truth value of (x1 <= x2) element-wise
<code>not_equal(x1, x2, /[, out, where, casting, ...])</code>	Return (x1 != x2) element-wise
<code>equal(x1, x2, /[, out, where, casting, ...])</code>	Return (x1 == x2) element-wise
<code>logical_and(x1, x2, /[, out, where, ...])</code>	Compute the truth value of x1 AND x2 element-wise
<code>logical_or(x1, x2, /[, out, where, casting, ...])</code>	Compute the truth value of x1 OR x2 element-wise
<code>logical_xor(x1, x2, /[, out, where, ...])</code>	Compute the truth value of x1 XOR x2, element-wise
<code>logical_not(x, /[, out, where, casting, ...])</code>	Compute the truth value of NOT x element-wise
<code>maximum(x1, x2, /[, out, where, casting, ...])</code>	Element-wise maximum of array elements
<code>minimum(x1, x2, /[, out, where, casting, ...])</code>	Element-wise minimum of array elements
<code>fmax(x1, x2, /[, out, where, casting, ...])</code>	Element-wise maximum of array elements
<code>fmin(x1, x2, /[, out, where, casting, ...])</code>	Element-wise minimum of array elements

Appendix B

Collaborative and Versioning Tools

...