Python Data Science Handbook

Chapter 1 : IPython: Beyond Normal Python

CHAPTER 1

IPython: Beyond Normal Python

There are many options for development environments for Python, and Iâ€™m often asked which one I use in my own work. My answer sometimes surprises people: my preferred environment is IPython plus a text editor (in my case, Emacs or Atom depending on my mood). IPython (short for Interactive Python) was started in 2001 by Fernando Perez as an enhanced Python interpreter, and has since grown into a project aiming to provide, in Perezâ€™s words, â€œTools for the entire lifecycle of research computing.â€• If Python is the engine of our data science task, you might think of IPyâ€• thon as the interactive control panel.
As well as being a useful interactive interface to Python, IPython also provides a number of useful syntactic additions to the language; weâ€™ll cover the most useful of these additions here. In addition, IPython is closely tied with the Jupyter project, which provides a browser-based notebook that is useful for development, collaboraâ€• tion, sharing, and even publication of data science results. The IPython notebook is actually a special case of the broader Jupyter notebook structure, which encompasses notebooks for Julia, R, and other programming languages. As an example of the useâ€• fulness of the notebook format, look no further than the page you are reading: the entire manuscript for this book was composed as a set of IPython notebooks.
IPython is about using Python effectively for interactive scientific and data-intensive computing. This chapter will start by stepping through some of the IPython features that are useful to the practice of data science, focusing especially on the syntax it offers beyond the standard features of Python. Next, we will go into a bit more depth on some of the more useful â€œmagic commandsâ€• that can speed up common tasks in creating and using data science code. Finally, we will touch on some of the features of the notebook that make it useful in understanding data and sharing results.

(fin del capitulo 1)

Chapter 2 : Introduction to NumPy

CHAPTER 2

Introduction to NumPy

This chapter, along with Chapter 3, outlines techniques for effectively loading, storâ€•
ing, and manipulating in-memory data in Python. The topic is very broad: datasets
can come from a wide range of sources and a wide range of formats, including collecâ€•
tions of documents, collections of images, collections of sound clips, collections of
numerical measurements, or nearly anything else. Despite this apparent heterogeneâ€•
ity, it will help us to think of all data fundamentally as arrays of numbers.
For example, imagesâ€”particularly digital imagesâ€”can be thought of as simply twodimensional arrays of numbers representing pixel brightness across the area. Sound
clips can be thought of as one-dimensional arrays of intensity versus time. Text can be
converted in various ways into numerical representations, perhaps binary digits repâ€•
resenting the frequency of certain words or pairs of words. No matter what the data
are, the first step in making them analyzable will be to transform them into arrays of
numbers. (We will discuss some specific examples of this process later in â€œFeature
Engineeringâ€• on page 375.)
For this reason, efficient storage and manipulation of numerical arrays is absolutely
fundamental to the process of doing data science. Weâ€™ll now take a look at the specialâ€•
ized tools that Python has for handling such numerical arrays: the NumPy package
and the Pandas package (discussed in Chapter 3.)
This chapter will cover NumPy in detail. NumPy (short for Numerical Python) proâ€•
vides an efficient interface to store and operate on dense data buffers. In some ways,
NumPy arrays are like Pythonâ€™s built-in list type, but NumPy arrays provide much
more efficient storage and data operations as the arrays grow larger in size. NumPy
arrays form the core of nearly the entire ecosystem of data science tools in Python, so
time spent learning to use NumPy effectively will be valuable no matter what aspect
of data science interests you.

(fin del capitulo 2)