

Python and Programming Basics

Practical Data Science

@jattenberg

May 2014

What is going on?

- Programming: translating a set of instructions into a format that a computer can understand and follow in order to complete a specific task.
- Programming Language: an artificial language designed to communicate instructions to a computer. While “coding” your goal will be to translate an abstract set of steps in your head into a programming language.

What is going on?

- MANY programming languages, have been developed, since FORTRAN in 1957, each with different properties
- We have already played with one language, unix shell scripting.
- Languages are distinguished along many dimensions, for us, one important dimension is compiled vs interpreted.
- Learning process analogous to learning a human language

Why bother?

- The practice of data science involves many interrelated but different activities:
 - accessing data
 - manipulating data
 - computing statistics about data
 - plotting/graphing/visualizing data
 - building predictive and explanatory models from data
 - evaluating those models on yet more data
 - integrating models into production systems, etc.

Why bother?

- It is possible to learn specialized tools that may accomplish one or two of these tasks, then figure how to put the pieces together somehow
- Or, it is possible to learn a more general tool (a robust programming language) that provides libraries to do these things. Python is an excellent choice.

Python

- For this class we will be using python (specifically, python 2.7)

Python

- A high-level programming language, favoring readability over flexibility and performance
- Interpreted rather than compiled: faster turn-around time
- Active user base: easy to find developers who know py and a rich set of online reading
- Incredible library support: as we'll explore, pre-built tools for many common tasks

Python

- **Object Oriented:** object oriented programming provides a concise representation of the objects being modeled. This paradigm emphasizes code reuse and abstraction of many details, enabling the development of very complex systems.
- **Functional:** functional programming naturally models the language of mathematical functions. Functional languages make it easy and efficient to translate an idea into a series of functions that are applied to some input. Commonly used in data science
- **Imperative Programming:** A series of steps are performed that change the state of the program and transform the input data. Most commonly used for simple scripts

Developing with Python

- There are three primary ways that we'll be working with python:
 - IPython: an interactive environment that executes python commands as you type them. Type “ipython” in the terminal to see!
 - Program (.py) files: self-contained programs or libraries. Often executed from the command line.

Developing with Python

- There are three primary ways that we'll be working with python:
 - IPython Notebooks: a web-based interactive programming environment. Try by typing “ipython notebook” in the terminal

IPython Notebooks

- Mixes code with results on a single web page
- Incorporate plots and markup to tell an entire story.
- A light weight web server starts up to host the notebooks
- Great for communicating analyses with others: show how you did your work!

IPython Notebooks

- Stored on disk as a “.ipynb” file.
- To restart where you left off, navigate the terminal to where your .ipynb file is, then run “ipython notebook”
- While running ipython notebook, code is executable and modifiable.
- commit your .ipynb files to version, and get “raw content”
- Use nbviewer to create public, static web pages from .ipynb files

IPython Notebooks

- A good tool for teaching python programming and for sharing your assignments!
- Note: unix commands can be executed in the ipython code windows by preceding with !: (!ls -lths to see all files in the folder you're running in)

Python Basics

- One command per line*
- *To stretch a single command over multiple lines, use “\ [return]”
- **It is possible to put multiple per like, separated by “;”
- Code blocks- pieces of code that are grouped together, are separated by a common indentation
- Unless there is some redirection (eg, a function call), code operates from top to bottom.

Let's dive in to python!

- Live IPython Notebook (or [http://nbviewer.ipython.org/github/jattenberg/PDS-Spring-2014/blob/master/ipython_notebooks/Python%20Basics%20\(Practical%20Data%20Science\).ipynb](http://nbviewer.ipython.org/github/jattenberg/PDS-Spring-2014/blob/master/ipython_notebooks/Python%20Basics%20(Practical%20Data%20Science).ipynb))
- Caveat: I am not a python expert, but I am a good programmer
- Caveat: There are often more than 1 way to things, what is presented here is a sample.