Desarrollo de Software en Astrofísica

Sala 763 Pabellón Forma,
Profesor Sebastián Pérez
23 de mayo 2023

Construcción de paquetes instalables en Python

En esta lección, hablaremos sobre cómo tomar código y convertirlo en algo que se pueda instalar en nuestro sistema a través de pip, y import en cualquier parte de nuestro sistema dentro de Python.

Por qué escribir un paquete?

A package, en abstracto, es solo código. La mayor parte del código que escribimos (en particular, el código de investigación) no constituye un paquete. Esto se debe a que nuestro código de investigación está escrito para manejar un conjunto de comandos hiperespecíficos para un proyecto único.

Esto no significa que no debamos hacer que nuestro código de investigación sea de acceso público. Es bueno para todos los involucrados y para la ciencia.

Un paquete, mientras tanto, es lo que sucede cuando descubres que algún fragmento de código que escribiste es aplicable de manera más general, ya sea para ti mismo más tarde o para otros.

Ejemplos de paquetes:

Fácil e intuitivo de usar!

emcee

emcee is an MIT licensed pure-Python implementation of Goodman & Weare's Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler and these pages will show you how to use it.

This documentation won't teach you too much about MCMC but there are a lot of resources available for that (try this one). We also published a paper explaining the emcee algorithm and implementation in detail.

emcee has been used in quite a few projects in the astrophysical literature and it is being actively developed on GitHub.

GitHub dfm/emcee | Tests passing | license MIT | arXiv 1202.3665 | coverage 96%

Basic Usage

If you wanted to draw samples from a 5 dimensional Gaussian, you would do something like:

```
import numpy as np
import emcee

def log_prob(x, ivar):
    return -0.5 * np.sum(ivar * x ** 2)

ndim, nwalkers = 5, 100
ivar = 1. / np.random.rand(ndim)
p0 = np.random.randn(nwalkers, ndim)

sampler = emcee.EnsembleSampler(nwalkers, ndim, log_prob, args=[ivar])
sampler.run_mcmc(p0, 10000)
```

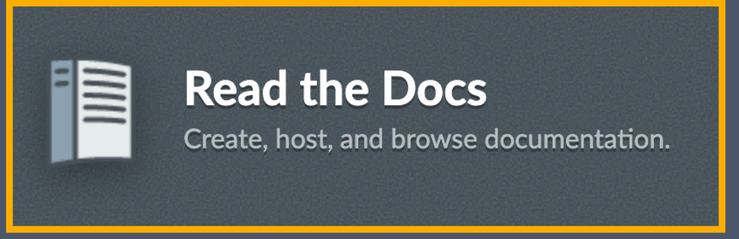
Una función como implot() para mostrar rápidamente imágenes astronómicas.

En lugar de copiar y pegar la función cada vez (que consume mucho tiempo y es propenso a las pesadillas del "control de versiones"), simplemente es mejor poder instalarla.

Por qué escribir un paquete?

El código bien mantenido, bien documentado y fácil de usar que está alojado en GitHub es una bandera verde distintiva para cualquiera que busque contratar a un científico de datos o informático.

Diseño de un paquete



Cosas a tener en cuenta al crear un paquete de código para que otros lo instalen y usen:

- Limitar las dependencias tanto como sea posible, idealmente a las más extendidas como `numpy`. Si usas una función de un modulo especializado, ¿puedes implementar esa única función ustedes mismxs?

- Tratar de hacer que las dependencias extrañas sean opcionales si es posible, teniendo bloques para manejar, ya sea que el usuario los tenga o no. `try: except`

- Intente que la instalación sea tan fácil como `pip install packagname` o `git clone package_address; cd packagename; pip install .`

- Tener buena documentación; tener un sitio `readthedocs` si es posible

- Mantener un registro de cambios en un lugar público (Github!), especialmente para cualquier cambio importante en la estructura o el uso del código.

- En los documentos, informar en qué versiones de python y numpy, etc. su código se probó más recientemente (e intentar mantenerlo actualizado a las últimas versiones)

- Si es posible (esto es elegante), tener un conjunto de pruebas que se ejecuten continuamente cuando se envíen cambios a su base de código, para verificar que nada se rompa.

Estructura de un paquete

Es hora de lo esencial. ¿Cómo hacemos realmente para hacer un paquete?
Más allá del código básico de Python, un paquete tiene tres componentes clave:

- Estructura (manejada a través de carpetas y archivos dentro de carpetas)
- `__init__.py` que definen el comportamiento de importación
- un `setup.py` que le dice a las bibliotecas de instalación estándar cómo instalar el paquete en una computadora.