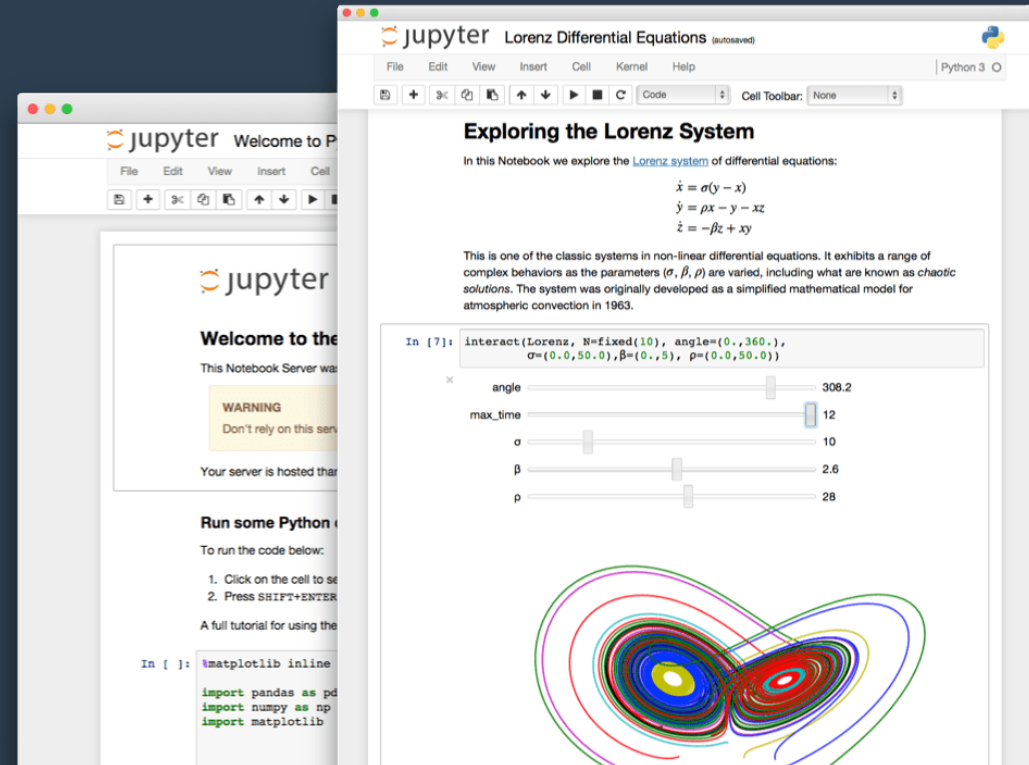


Curso Introductorio al Modelado de Datos en Python

Primer Encuentro

11/06/19



Lic. Alan Taitz

Programa del Curso

Encuentro 1: Introducción a Python

Encuentro 2: Python Científico.

Encuentro 3: Modelado de Datos I

Encuentro 4: Modelado de Datos II

Programa del Curso

Encuentro 1: Introducción a Python

a. Anaconda

Spyder, Jupyter Notebook, librerías

b. Tipos de Variables

Números, Listas, Cadenas, Diccionarios y Tuplas.

c. Funciones

d. Estructuras de Control:

If, For y While

e. Objetos

Clases, Herencia, Clases Abstractas

f. *Testing, Debugging, Assertion*

Encuentro 2: Python Científico.

Encuentro 3: Modelado de Datos I

Encuentro 4: Modelado de Datos II

Qué es Anaconda

- **Anaconda** es una distribución de los lenguajes **Python** y **R**, muy utilizada en **ciencia de datos** y **aprendizaje automático**.
- Está orientado a simplificar el despliegue y administración de los **paquetes de software científico**.
- Involucra el **procesamiento** de grandes volúmenes de información, **análisis predictivo** y **cálculos científicos**.
- Permite coordinar paquetes para que sean **compatibles**. Hace muy sencillo instalar, correr y actualizar software de ciencia de datos como **Scikit-learn**, **TensorFlow** y **SciPy**.

Anaconda



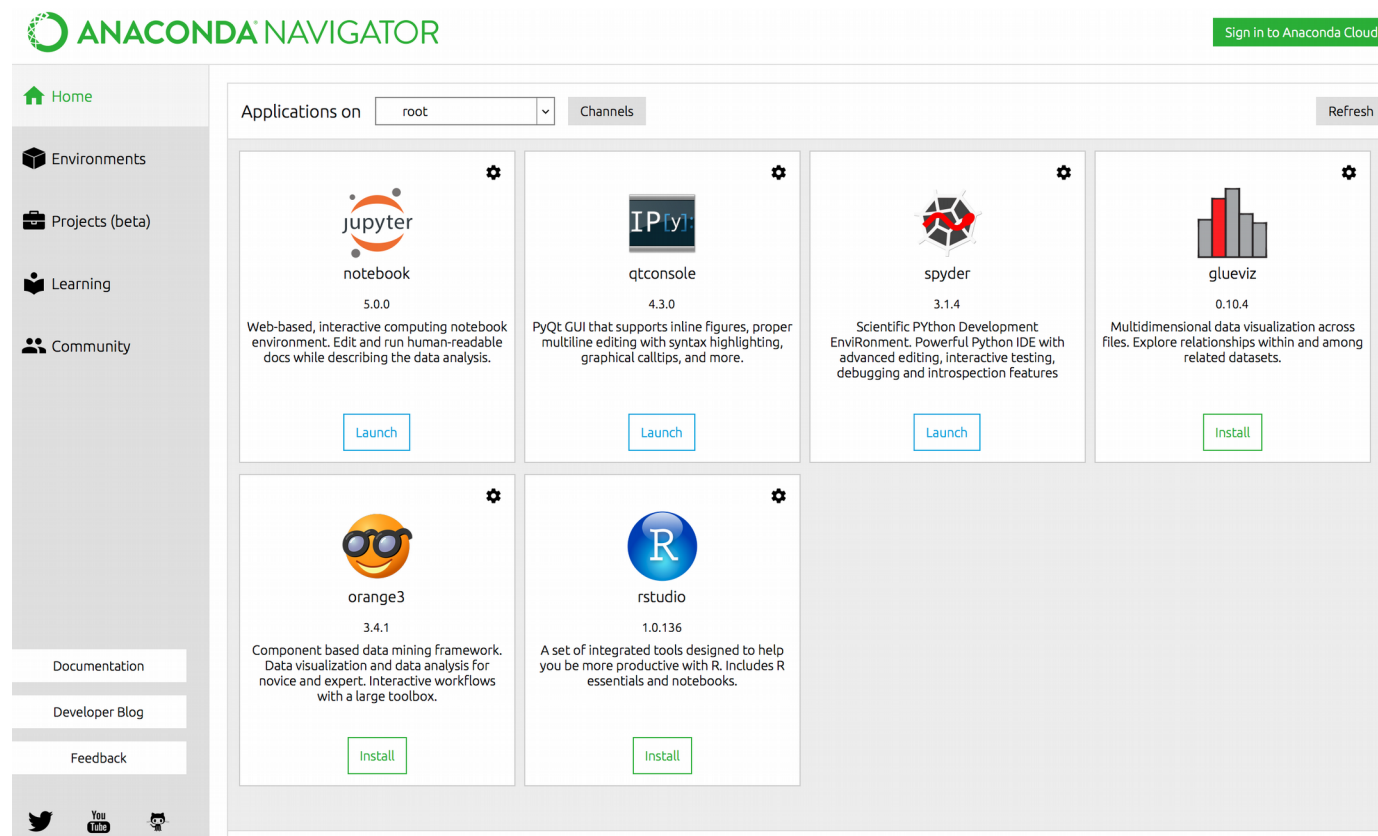
<https://www.anaconda.com/distribution/>



Windows

Install Wizard

ANACONDA NAVIGATOR



Entornos

- La gran ventaja de anaconda es que permite crear **entornos**
- **Entornos** son unidades separadas donde puedo tener distintas versiones de python corriendp con ciertos donde los paquetes instalados
- Anaconda los gestiona de manera tal que dentro de cada entorno los paquetes estén **coordinados**: versiones compatibles, evita la superposición de paquetes, etc
- Ideal para colaborar, entregar aplicaciones a un cliente o para moverse entre computadoras uno mismo: se pueden **exportar**

Porqué Python

- **Fácil** de aprender
- **Escalabilidad**
Flexible, rápido y versátil para escalar: hasta YouTube migró a Python.
- Variedad de **librerías**
Pandas, StatsModels, Numpy, SciPy y miles más que resuelven problemas alta complejidad de manera sencilla.
- **Comunidad enorme y activa**
Voluntarios creando y modificando librerías continuamente de manera gratuita
Amplias soluciones a problemas con tan solo googlear consultas.
- **Gráficos y Visualización**
Paquetes de visualización de datos que ofrecen una variedad de gráficos asombrosa.

Spyder



Integrated Development Environment (IDE) multiplataforma de código abierto para programación científica en el lenguaje Python.

The screenshot displays the Spyder Python IDE interface. The main window is titled "Spyder (Python 3.4)". The top menu bar includes File, Edit, Search, Source, Run, Debug, Consoles, Tools, View, and Help. The left sidebar shows the Editor with a file named "interpolation.py" open. The code in the editor is as follows:

```
4 From the SciPy Cookbook
5 """
6
7 from numpy import arange, cos, linspace, pi, sin, random
8 from scipy.interpolate import splprep, splev
9
10 # make ascending spiral in 3-space
11 t=linspace(0,1.75*2*pi,100)
12
13 x = sin(t)
14 y = cos(t)
15 z = t
16
17 # %% add noise
18 x+= random.normal(scale=0.1, size=x.shape)
19 y+= random.normal(scale=0.1, size=y.shape)
20 z+= random.normal(scale=0.1, size=z.shape)
21
22 # %% spline parameters
23 s=3.0 # smoothness parameter
24 k=2 # spline order
25 nest=-1 # estimate of number of knots needed (-1 = maximal,
26
27 # %% find the knot points
28 tckp,u = splprep([x,y,z],s=s,k=k,nest=-1)
29
30 # %% evaluate spline, including interpolated points
31 xnew,ynew,znew = splev(linspace(0,1,400),tckp)
32
33 import pylab
```

The right sidebar contains the Object inspector, which shows the "mean" function. The definition is: `mean(a, axis=None, dtype=None, out=None, keepdims=False)`. The type is: `Function of numpy.core.fromnumeric module`. Below the definition, it states: "Compute the arithmetic mean along the specified axis. Returns the average of the array elements. The average is".

The bottom right pane shows the IPython console. It displays the command: `runfile('/tmp/interpolation.py', wdir='/tmp')`. Below the command, there are two plots. The left plot shows a 3D scatter plot of data points (blue dots) and a fitted spline (red line). The right plot shows a 2D scatter plot of data points (blue dots) and a fitted spline (red line).

The status bar at the bottom indicates: Permissions: RN, End-of-lines: LF, Encoding: UTF-8, Line: 18, Column: 43, Memory: 86 %.

Jupyter Notebook



Es una **aplicación web** de **código abierto** que permite crear y compartir **documentos**

- Se está convirtiendo en el **standard** en el Análisis de Datos.
- **Comunidad** muy activa
- Ampliamente utilizado por los **principales laboratorios** del mundo.
- Recomendado por **Nature** en un artículo reciente.

