

M2.855 Modelos avanzados de minería de datos

Preparación del entorno: Instalación de Python y IPython

Preparación del entorno de prácticas

Este documento nos guiará en la instalación de todo el software necesario para poder desarrollar las actividades prácticas de esta asignatura. Como se verá más adelante en este mismo documento, el entorno se basa en el lenguaje de programación Python y las librerías necesarias para poder manipular, analizar y visualizar datos.

Para la realización de las actividades de esta asignatura es necesario que cada estudiante instale y configure en su ordenador el lenguaje Python, las librerías necesarias y (opcionalmente) un entorno de desarrollo integrado (IDE).

Instalación de Python y IPython

Para esta y las siguientes actividades, vamos a utilizar Python 3.7, que está disponible para sistemas Linux, Mac OS X y Windows en el sitio web oficial de Python:

<https://www.python.org/downloads/>

Hay dos versiones principales para elegir, Python 3.x y Python 2.x. La versión 2.x, aunque aún muy utilizada se mantiene por motivos de compatibilidad con software existente. En nuestro caso optamos por la última versión disponible de la serie 3.x.

Después de la instalación de Python, podemos comprobar su correcto funcionamiento accediendo a la terminal de Python como se muestra a continuación:

```
$ python3
Python 3.7.0 (default, Jun 28 2018, 08:04:48)
[GCC 4.8.2] on Linux
Type ``help'', ``copyright'', ``credits'' or ``license'' for
more information.
>>> print(`Hello, world!')
Hello, world!
>>> exit()
```

IPython es una plataforma para el desarrollo de Python que contiene una serie de herramientas y entornos para ejecutar Python y contiene, además, funciones adicionales al intérprete estándar. Entre otras, esta plataforma contiene el potente IPython Notebook¹, que te permite escribir programas en un navegador web. También formatea tu código, muestra la salida y te permite anotar tus *scripts*. Es una herramienta útil e interesante para explorar conjuntos de datos y realizar análisis más

¹ Más información en <https://ipython.org/notebook.html>

o menos sencillos. En el caso de ejecutar scripts complejos que puedan llevar horas de ejecución, aconsejamos utilizar el intérprete estándar de Python.

Para instalar IPython en nuestro sistema, debemos escribir lo siguiente en la línea de comandos del sistema operativo (no en el intérprete de Python):

```
$ pip3 install ipython[all]
```

Se necesitan privilegios de administrador para instalarlo en el sistema y que sea accesible para todos los usuarios. Si no deseamos (o no podemos) realizar cambios en todo el sistema, podemos instalarlo sólo para el usuario actual ejecutando el siguiente comando:

```
$ pip3 install --user ipython[all]
```

Esto instalará el paquete IPython en una ubicación específica del usuario; podremos utilizarla, pero nadie más en nuestro equipo podrá ejecutarlo.

Para una información detallada sobre la instalación, podemos referirnos a la documentación oficial² para obtener instrucciones de instalación más detalladas.

Una vez la plataforma ha sido instalada en nuestro sistema, podemos ejecutarla mediante la orden:

```
$ ipython3 notebook
```

Esto hará dos cosas. En primer lugar, creará una instancia de IPython Notebook que se ejecutará en el intérprete de comandos que acaba de utilizar. En segundo lugar, lanzará su navegador web y se conectará a esta instancia, lo que le permitirá crear un nuevo proyecto de trabajo, como se puede ver en la figura 1.

² Disponible en <http://ipython.org/install.html>

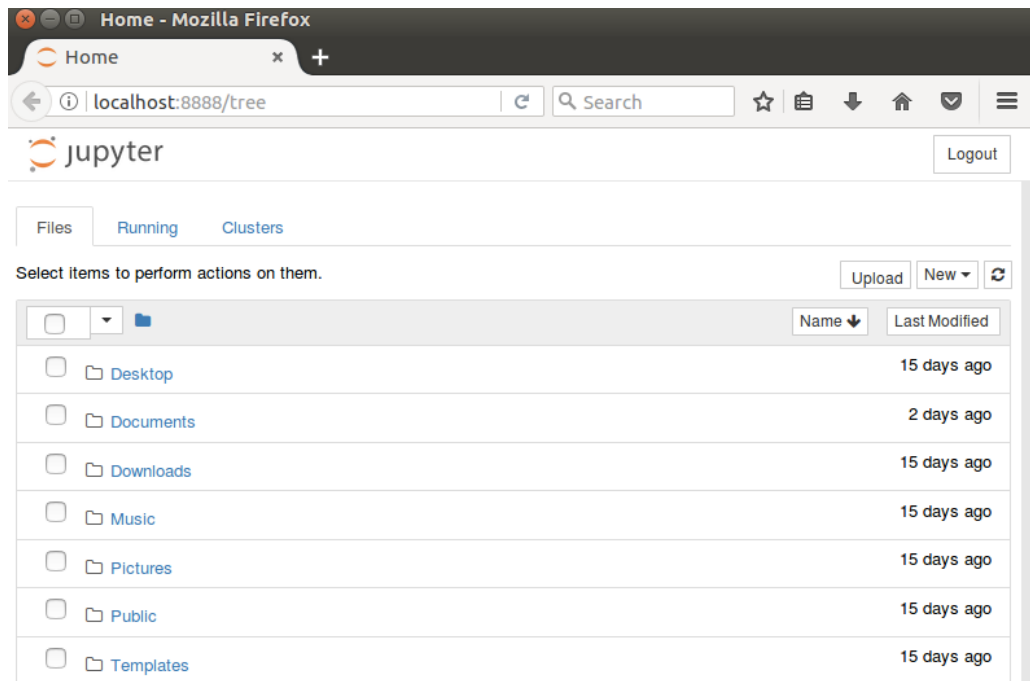


Figura 1. Interfície web de IPython notebook (Jupyter notebook)

Alternativamente, se puede instalar una distribución de Python llamada Anaconda³. Esta distribución incluye muchas de las librerías utilizadas en ciencia de datos y *machine learning*, además de la interfaz web basada en IPython y un entorno integrado para la gestión de los servicios y paquetes.

Instalación de las librerías

Como hemos comentado anteriormente, Python incluye un programa llamado *pip*, que es un paquete extremadamente útil que nos ayudará a instalar nuevas librerías en nuestro sistema. Se puede verificar que *pip* está instalado en nuestro sistema ejecutando el siguiente comando:

```
$ pip3 freeze
```

Que nos indica que paquetes tenemos instalados en nuestro sistema.

La instalación de cualquier librería es un proceso muy sencillo empleando la herramienta *pip*. Para instalar cualquier librería sólo deberemos indicar la opción “install” y el nombre de la librería que deseamos instalar.

Por ejemplo, para instalar la librería “numpy” ejecutaremos el siguiente comando desde la línea de órdenes del sistema operativo:

³ Disponible en <https://www.anaconda.com/distribution/>

```
$ pip3 install numpy
```

Las librerías básicas y necesarias para realizar esta actividad se detallan a continuación:

- numpy
- pandas
- matplotlib
- scikit-learn

Entorno de desarrollo integrado (IDE)

Podemos utilizar cualquier editor de texto para codificar los programas en Python. Por lo tanto, este paso es opcional, y depende mucho de las preferencias de cada usuario.

Una primera opción son los entornos “ligeros”, como editores de texto estándar que reconocen la sintaxis de Python y permiten indentar y colorear el código (lo cual es de gran ayuda). En este sentido se pueden encontrar muchísimos editores disponibles y gratuitos (algunos) con una simple búsqueda por Internet. No haremos recomendaciones en este sentido.

Por otro lado, existen entornos de desarrollo integrados que van mucho más allá de la simple edición de código. Estos IDE permiten, además de facilitar la codificación, el uso de *debuggers* y otras herramientas que facilitan el desarrollo de proyectos y código. Entre otros, existe el entorno comercial PyCharm⁴, que dispone de una versión *community* y una versión *professional* (que incorpora funciones adicionales y puede ser utilizada de forma gratuita por estudiantes e investigadores del mundo académico).

Otra opción interesante es Spyder⁵, que ofrece una interfaz similar al conocido RStudio, pero para el lenguaje de programación Python. Además, en este caso se ofrece bajo una licencia de uso libre.

⁴ Más información en <https://www.jetbrains.com/pycharm/>

⁵ Más información en <https://github.com/spyder-ide/spyder>