

## Entornos de data science con Python

Bloque 1 – Guía de estudio

2015/16

---

### Introducción

Las bibliotecas fundamentales del stack de Python son los **arrays de NumPy** y los **DataFrames de pandas**. Los primeros son arrays homogéneos eficientes sobre los que se implementan **operaciones vectorizadas**, y los segundos son tablas preparadas para cargar datos externos y hacer análisis sobre ellas.

Un manejo eficaz de arrays de NumPy y DataFrames es fundamental para el resto de las tareas de data science con Python. La totalidad de las bibliotecas de SciPy utilizan NumPy como elemento base, y también muchas de ellas se integran con Pandas, por lo que se pueden considerar como el “Python científico básico”.

Un array de NumPy realmente puede ser multidimensional, aunque también tiene un tipo derivado matrix que tiene operaciones más convenientes como la multiplicación de matrices con operadores, y es estrictamente bidimensional.

### Objetivos específicos

- Saber utilizar los arrays de la biblioteca NumPy con una y dos dimensiones.
- Entender las diferencias entre los arrays de NumPy y las estructuras básicas de Python, en lo relativo a funcionalidad (operaciones vectorizadas) y a eficiencia.
- Saber manipular arrays y matrices de Numpy.

### Recursos

Además de la información que se puede encontrar en las páginas de NumPy (<http://www.numpy.org/>), los siguientes son recursos interesantes para reforzar el estudio:

- Un tutorial completo básico “desde cero” reciente de Eric Jones lo podéis encontrar aquí: <https://www.youtube.com/watch?v=1zmV8IZsHF4>
- Una visión general de NumPy se proporciona en el capítulo 4 de PFDA<sup>1</sup>.
- El Seminar for Applied Mathematics (SAM) tiene un tutorial de NumPy bastante interesante: [http://www.sam.math.ethz.ch/~raoulb/teaching/PythonTutorial/intro\\_numpy.html](http://www.sam.math.ethz.ch/~raoulb/teaching/PythonTutorial/intro_numpy.html)

---

<sup>1</sup> McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc.

- En “SciPy Lecture Notes” hay un capítulo interesante para repasar lo fundamental de NumPy, concretamente las secciones 1.3.1 y 1.3.2 de este documento: <http://www.scipy-lectures.org/intro/numpy/index.html>
- Un recurso interesante para volver a practicar NumPy es el documento “An introduction to NumPy and SciPy” del curso CHE210D de la UCSB, que puede obtenerse aquí: <http://engr.ucsb.edu/~shell/che210d/assignments.html>

### Actividades

- Trabajar alguno de los recursos mencionados u otros similares.
- Ejercicios prácticos propuestos.