

# Trabajando con NumPy y Pandas: Prueba de evaluación

ASIGNATURA: Entornos de Data Science en Python  
Última actualización: 30 de enero de 2015

## Objetivos

- Adquirir práctica en el manejo de la herramienta IPython Notebook.
- Saber crear arrays de una y dos dimensiones con NumPy, manipularlos (indexing/slicing) y utilizar funciones vectorizadas sobre ellos.
- Saber manipular datos tabulares con Pandas, incluyendo agrupaciones.
- Ser capaces de crear visualizaciones básicas con matplotlib.

## Criterio de evaluación

Esta es una práctica básica de manejo de las bibliotecas numéricas NumPy y Pandas. Se evaluará simplemente el ser capaces de realizar las áreas solicitadas dentro de un IPython Notebook.

Esta prueba tiene un peso de un 20% de la evaluación del módulo

## 1. NumPy en una dimensión

- Crea un array `a` de NumPy con números aleatorios siguiendo una distribución logística.
- Crea un segundo array `b` multiplicando el anterior por una constante.
- Crea otro array adicional `c` del mismo tamaño con aleatorios de una distribución normal.
- Comprueba e interpreta un coeficiente de correlación de Pearson (biblioteca `scipy.stats`) entre los arrays creados, interpretando el resultado.
- Dibuja las distribuciones (histogramas) de los tres arrays.
- Crea un nuevo array `X` con la mitad de los elementos de `a` y la mitad de los elementos de `c`. Asegúrate de que el array `X` es un nuevo array en memoria, no una vista sobre slices de los anteriores.
- Ordena el array anterior. Investiga si el array ordenado es una copia o una vista.
- Busca una función en las bibliotecas de NumPy que nos permita generar una permutación aleatoria de un array. Crea un array de enteros en un espacio lineal (`linspace`) y utiliza esa función para generar una permutación.
- Investiga las funciones `loadtxt` y `savetxt` y usalas para guardar el array anterior.

- Crea un array con números aleatorios enteros entre 1 y 10. Utilizando indexación con condiciones, extrae los elementos que son mayores que cinco en un nuevo array.

## 2. NumPy en dos dimensiones

- Crea una matriz de números reales con NumPy con números aleatorios con la forma (6, 4).
- Establece a cero todos los elementos de la tercera columna.
- Crea un slice con la submatriz que cubre los elementos que están en las dos primeras columnas y las dos primeras filas (una matriz de 2x2). Crea una copia de ese slice en una nueva matriz. Comprueba que no es una vista, sino una copia nueva en memoria.
- Crea un array Z de 50 elementos aleatorios entre 10 y 20 y rediménsionalo para que sea una matriz de 10 filas y cinco columnas.
- Extrae a un array los elementos de la matriz Z que están entre 15 y 17, y pon esos elementos a cero en Z.
- Obtén la media por filas y por columnas de los elementos del array Z.
- Obtén la suma acumulada (tienes que buscar la función adecuada de NumPy) de los elementos de Z y explica los resultados.

## 3. Pandas

- Crea un DataFrame nuevo que tenga valores nulos (NaN). Utiliza fillna() para reemplazar esos valores por otros no nulos. Investiga si esa función hace una copia o modifica el DataFrame original.
- Crea un DataFrame vacío y añade campos y datos de ejemplo para llevar una agenda de teléfonos, que incluya la ciudad en la que vive cada contacto.
- Ordena tu agenda por apellido.
- Agrupa los contactos por ciudad, y cuenta cuántos contactos tienes en cada ciudad.
- Obtén los contactos de tu agenda que viven en Barcelona.
- Crea un DataFrame nuevo para llevar tu calendario. Lo que queremos guardar son citas: fecha, hora, persona con la que has quedado, motivo y lugar. Rellénalo con datos (para que tengan sentido los siguientes ejercicios), algunos de ellos que vengan de la creación del DataFrame a partir de un diccionario de Python.
- Haz que el DataFrame tenga como índice la fecha.
- Obtén las citas de los meses de junio y agosto.
- Crea una función en Python que devuelva una lista de las citas para el día que le pases como parámetro.
- Elimina todas las citas de marzo del DataFrame.
- Obtén una tabla con los contactos y el número de citas que tienes en la agenda con cada uno de ellos.