

Introducción al lenguaje de programación Python

Presentación

Curso: 2018 - 2019

PROFESOR

Guillermo González Sánchez

Científico de datos en Strategy Big Data

Licenciado en Matemáticas

Semana On-line (16h)

TEMA 1. Conceptos generales y Python científico

- El proceso de la minería de datos.
- Concepto de aprendizaje automático.
- Limpieza y transformación de datos.
- Python en ciencia de datos:
 - Numpy
 - Pandas
 - Visualización (matplotlib, seaborn)
- Representación de datos.

Actividad 1. Análisis de un Dataset

Se realiza en un notebook y se rellena online en un cuestionario para evaluar el trabajo.

Corresponde al 50% de la nota final.

Entrega antes del Jueves 25 de octubre de 2018 antes de las 16 horas

Abrimos las transparencias de las tres sesiones del tema 1.

Es mucho material pero vamos a comentar qué notebooks son los fundamentales.

- **Sesión 1: Anaconda, elementos Python y Numpy**
 - Todos
 - Hay dos actividades (no evaluables) que aparecen resueltas
- **Sesión 2: Pandas**
 - Los fundamentales son hasta 4a
 - Hay una actividad no evaluable básica.
 - Muy importante el notebook 4a para la actividad evaluable que haremos.
 - Si se quiere comprender todo mejor se puede leer el libro **Python for Data Analysis** (caps 5, 7 y 9 fundamentalmente)
<http://www3.canisius.edu/~yany/python/Python4DataAnalysis.pdf>
- **Sesión 3: Matplotlib, y Seaborn**
 - Los notebooks más importantes son el 8 y el 9
 - Interesante mirar las actividades (no evaluables) para entender el proceso de representación y qué podemos averiguar.

¿ Linux, que hacía antes de conocerte ?

- En ciencia de datos todo se materializa programando, tanto el producto, como las representaciones, como los reportes de calidad con alto grado de ajuste.
- Las mejores librerías con algoritmos, visualizaciones y estadística están en R y Python y son de **código libre**.
- Los programadores que desarrollan estas librerías en código libre, lo hacen en su mayoría en entornos **LINUX**. Por tanto, los rendimientos son superiores en este sistema operativo.



¿ Soy principiante en Linux?

ubuntu[®]

- Si bien al principio la curva de aprendizaje es dura de llevar, cuando os hagáis con el terminal **bash**, nunca querréis dejar de tenerlo a vuestro lado.

Por tanto: imprescindible tener partición de disco o disco completo con SO Linux.

Usaremos anaconda para la versión de Python 3.5 o 3.6

<https://www.continuum.io/downloads>



- Vemos cómo se instala en los diferentes sistemas operativos
- Vemos cómo se abre un notebook en Windows y en Linux
- Comentamos el IDE Jupyter Notebook y abrimos unos pocos
- También se puede acceder en AWS en la máquina virtual de la que disponen los alumnos