

Librerías de Python para tratamiento de datos

Numpy

Para trabajar con arrays, vectores. Soporte para arrays multidimensionales

<https://www.numpy.org>

Scipy

Orientada a matemáticas, ciencias e ingeniería

<https://www.scipy.org>

Pandas

Manipulación y análisis de datos. Ofrece estructuras de datos para la manipulación de tablas, series temporales, indexación, agrupación, lectura de datos(csv,sql,excel)

<http://pandas.pydata.org>

NumPy

Es una librería que incluye la funcionalidad necesaria para poder llevar a cabo computación científica con Python. Permite manejar vectores y matrices, así como realizar operaciones de álgebra lineal, transformaciones y manejo de números aleatorios. Esta librería es fundamental para poder computar el algoritmo del sistema de recomendación.

Comenzaremos nuestra serie Python for Data Scientists con NumPy , abreviatura de Numerical Python, que es el paquete principal para la computación científica en Python. Para los datos numéricos, los arrays NumPy son una forma mucho más eficiente de almacenar y manipular datos que las otras estructuras de datos integradas de Python.

Estas son algunas de las cosas que ofrece:

- Un objeto de matriz multidimensional rápido y eficiente ndarray
- Funciones para realizar arrays de cálculos de elementos
- Herramientas para leer y escribir conjuntos de datos basados en matriz en disco
- Operaciones de álgebra lineal, transformada de Fourier y generación de números aleatorios
- Herramientas para integrar la conexión de C, C ++ y código Fortran a Python

Instalación

Dado que todo el mundo utiliza Python para diferentes aplicaciones, no existe una única solución para configurar Python y los paquetes adicionales necesarios.

Personalmente recomiendo usar una de las siguientes distribuciones Python base:

- **Enthought Python Distribution** : una distribución Python orientada a la ciencia de datos . Esto incluye Canopy Express , una distribución científica de base gratuita (con NumPy, SciPy, matplotlib, Chaco e IPython) y Canopy Full, una suite completa de más de 300 paquetes científicos en muchos dominios.
- **Python (x, y)** : Una distribución gratuita de Python orientada a la ciencia de datos para Windows.
- **pip install numpy**

Ndarray: un objeto de matriz multidimensional

Una de las características clave de NumPy es su objeto de matriz N-dimensional, o ndarray, que es una estructura rápida y flexible para grandes conjuntos de datos en Python. Las matrices permiten realizar operaciones matemáticas en bloques completos de datos utilizando una sintaxis similar a las operaciones equivalentes entre elementos escalares. Esto es importante porque permiten expresar las operaciones por lotes en los datos sin necesidad de escribir los bucles. Esto se suele denominar vectorización.

<https://docs.scipy.org/doc/numpy/reference/generated/numpy.ndarray.html>

SciPy

SciPy está construido sobre NumPy y proporciona muchas rutinas numéricas fáciles de usar y eficientes que abordan una serie de diferentes dominios de problemas estándar en computación científica como integración, solucionadores de sistemas lineales dispersos y diferenciados, optimizadores y algoritmos de búsqueda de raíces...

Instalación

```
pip install scipy
```

Pandas

Es una librería de Python que permite manejar fácilmente estructuras de datos y herramientas de análisis de datos. Permite manejar diferentes fuentes de datos: datos tabulados, series temporales, matrices de datos, datos estadísticos.

Para instalar dicho módulo es necesario Python 2.7 o superior y la librería Numpy(3.5.5) versión 1.7 o superior.

Se utilizará dicha librería para manejar el conjunto de datos escogido para entrenar el modelo.