

Python na Ciência



Raony Guimarães
8º Encontro de Python de Minas Gerais
21/03/2015

\$whoami

- Estudante de Doutorado em Bioinformática pela UFMG.
- Bacharel em Informática Biomédica pela USP
- Especialista em análise de exomas humanos para diagnóstico de doenças mendeliana.

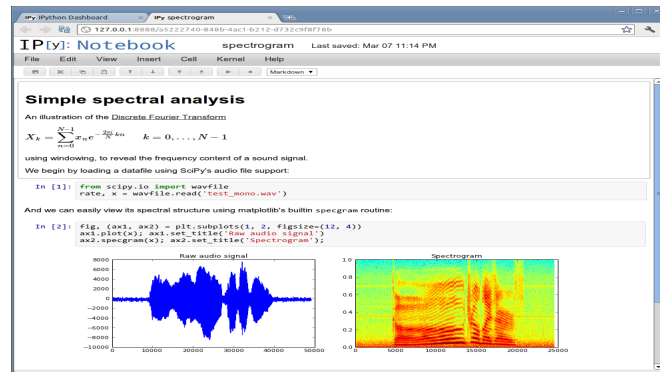
**Para que o python é
utilizado na ciência?**

Quais são as necessidades de um cientista?

- Obter os dados
- Manipular e processar os dados
- Visualizar os resultados
- Comunicar os resultados: produzir figuras para relatórios, publicações e apresentações

ipython-notebook

O IPython Notebook é um ambiente computacional interativo baseado em web onde você pode combinar, execução do código, texto, matemática, plots e rich media em um único documento.



Numpy



NumPy é o pacote fundamental para computação científica utilizando Python.

- Um poderoso objeto de array N-dimensional
- Funções sofisticadas
- Ferramentas para integrar código C/C++ e Fortran
- Álgebra linear, transformada de Fourier, e capacidade de gerar números aleatórios

matplotlib



Matplotlib é o pacote de Python mais utilizado para gráficos 2D. Ele fornece uma maneira muito rápida de visualizar dados de Python e figuras de qualidade para publicação em vários formatos.

Scipy



- SciPy é uma coleção de algoritmos matemáticos e funções de conveniência construídas sobre a extensão Numpy de Python.
- Com SciPy, uma sessão interativa de Python torna-se um ambiente de processamento de dados e sistema de prototipagem rivalizando com sistemas como MATLAB, IDL, Octave, R-Lab, e SciLab.

scipy is composed of task-specific sub-modules:

<code>scipy.cluster</code>	Vector quantization / Kmeans
<code>scipy.constants</code>	Physical and mathematical constants
<code>scipy.fftpack</code>	Fourier transform
<code>scipy.integrate</code>	Integration routines
<code>scipy.interpolate</code>	Interpolation
<code>scipy.io</code>	Data input and output
<code>scipy.linalg</code>	Linear algebra routines
<code>scipy.ndimage</code>	n-dimensional image package
<code>scipy.odr</code>	Orthogonal distance regression
<code>scipy.optimize</code>	Optimization
<code>scipy.signal</code>	Signal processing
<code>scipy.sparse</code>	Sparse matrices
<code>scipy.spatial</code>	Spatial data structures and algorithms
<code>scipy.special</code>	Any special mathematical functions
<code>scipy.stats</code>	Statistics

Scikit-Learn

- Ferramentas simples e eficientes para a mineração e análise de dados
- Acessíveis a todos, e reutilizáveis em vários contextos
- Construído sobre NumPy, SciPy e matplotlib
- Open source, comercialmente utilizáveis - licença BSD

Classification

Identifying to which set of categories a new observation belong to.

Applications: Spam detection, Image recognition.

Algorithms: *SVM, nearest neighbors, random forest, ...* — Examples

Regression

Predicting a continuous value for a new example.

Applications: Drug response, Stock prices.

Algorithms: *SVR, ridge regression, Lasso, ...* — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: *k-Means, spectral clustering, mean-shift, ...* — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency

Algorithms: *PCA, feature selection, non-negative matrix factorization.* — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning

Modules: *grid search, cross validation, metrics.* — Examples

Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.

Modules: *preprocessing, feature extraction.* — Examples

Aplicações

- **Programação e Ciências da Computação**
- **Estatísticas, Aprendizado de Máquinas e Data Science**
- **Matemática, Física, Química, Biologia**
- **Ciências da Terra e dados Geo-Espaciais**
- **A visualização de dados e plotagem**
- **Processamento de Sinal e Som**
- **Processamento da Língua Natural**
- **Lingüística e Text Mining**

It's Show Time!

Obrigado!

@raonyguimaraes

<http://raonyguimaraes.com>

raonyguimaraes@gmail.com

```
random.choice(['pergunta', 'dúvida', 'sugestão'])
```

