

Como instalar o Python

- **Como instalar o Python e as bibliotecas que vamos usar?**

Uma maneira fácil de instalar o Python é através de uma distribuição, como a Anaconda. É gratuita e já vem com a maior parte dos pacotes que vamos usar (numpy, pandas, seaborn, etc. O tensorflow precisa ser instalado separadamente). Pra instalar é só escolher o sistema operacional certo, baixar e seguir as instruções no site.

Instalar bibliotecas pelo Anaconda se resume (no Linux, pelo menos) ao comando “conda install package”, onde “package” é o nome da biblioteca. Por exemplo, “conda install tensorflow”. Pelo Python puro, em geral usamos o pip, “pip install package”, por exemplo, “pip install tensorflow”.

- **Algumas referências sobre Python e ciência de dados**

Saber bem alguma linguagem de programação é necessário. Em ciência de dados a linguagem que se tornou dominante é o Python. Existem diversos cursos online de Python e muito material disponível na internet. Alguns livros que gosto bastante são os do Mark Lutz (Learning Python, o mais inicial dele, salvo engano). Um outro já mais avançado bem legal é o do Luciano Ramalho, Fluent Python (também disponível em português).

Curso Python - 2020

Usina Jr



- Uma outra linguagem muito usada em ciência de dados pra acessar bancos de dados é algum dialeto de SQL (Structured Query Language). Também com muito material disponível na internet, como nos minicursos do Kaggle.
- Em português existe a comunidade do Data Hackers de ciência de dados, com posts, canal do Slack, podcasts, etc., onde há muito material útil, atual e relevante. Outro canal com bastante informação útil é o canal do Mario Filho no Youtube.
- Já em inglês a quantidade de material é bem mais vasta. Alguns livros interessantes com overviews da área incluem o Data science design manual do Skiena, Data science for business, do Provost & Fawcett (também disponível em português).
- Pra uma introdução aos algoritmos básicos (e outros nem tão básicos) usados em ciência de dados um livro que eu usei e gosto bastante é o Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, do Géron (salvo engano tem duas edições. A segunda é com o tensorflow atualizado e por isso recomendo essa. A primeira está disponível em português).
- Existem vários blogs com conteúdo útil, tanto avançado quanto básico. Alguns incluem o towardsdatascience, fivethirtyeight, dataelixir, etc.

Material produzido pelo palestrante

Link do notebook utilizado:

https://colab.research.google.com/drive/1fpjifCDLwjf1J2h6a76QEx_gRh9d3EA?usp=sharing

MINISTRADO POR: GABRIEL COZZELLA