

INTRODUÇÃO À CIÊNCIA DE DADOS

Biblioteca NumPy

BIBLIOTECA NUMPY



- NumPy é a abreviatura de Numerical Python.
- É um dos pacotes mais importantes para processamento numérico em Python. Matemática é o cerne do DS!
- A maioria dos pacotes de processamento com funcionalidades científicas utiliza objetos array do NumPy.
- Muito utilizado para array (listas) e também para array multidimensional com operações aritméticas rápidas, sem uso de laços, além de muitos recursos.
- UMA PAUSA: um profissional de DS precisa dominar o NumPy e principalmente as estruturas de dados do Python, arrays multidimensionais principalmente.

BIBLIOTECA NUMPY

- A biblioteca NumPy é um assunto muito amplo, seria possível fazer uma disciplina só disso!
- Vamos abordar aqui algumas características importantes da biblioteca, mas é muito importante recorrer ao livro texto, ao site da biblioteca e testar muito o uso das funcionalidades.
- Algumas funcionalidades importantes de DS que o NumPy ajuda a resolver:
 - operações rápidas em vetores para tratamento, limpeza, geração de subconjuntos, transformações e outros processamentos;
 - operações para gerar ordenações, unicidade e operações de conjuntos.

Continua... 

BIBLIOTECA NUMPY

- estatísticas descritivas, agregação e sintetização de dados.
- expressões lógicas condicionais na forma de expressão ao invés de usar laços com desvios (if, else...)
 - fundamental compreender isso para ser um ótimo profissional de DS. Se você iniciou programando em outra linguagem, prepare-se para uma mudança de paradigma.
 - além disso a diferença de performance é brutal (10 a 100x mais rápido).
- manipulação de dados em grupos (agregação, transformação e aplicação de funções).
- Entre muitas outras!

BIBLIOTECA NUMPY

- Um dos motivos que tornam a biblioteca NumPy tão importante é o fato dela ter sido projetada para ser muito eficaz ao lidar com arrays de dados muito grandes.
- É provável que mais adiante muitos prefiram usar a biblioteca Pandas (que veremos mais pra frente) para atividades com arrays multidimensionais, mas para algumas tarefas o NumPy é mais eficaz.

NDARRAY

- O objeto de array N-dimensional (ndarray), é um container rápido e flexível para grandes conjuntos de dados em Python.
- Permitem realizar operações matemáticas em blocos inteiros, com sintaxe semelhante a operações realizadas com dados escalares.
- Todos elementos em um ndarray são do mesmo tipo, por meio do atributo *shape* é possível saber a dimensão do objeto e o atributo *dtype* permite saber o tipo de dado do array.
- Há uma lista de atributos, funções e métodos disponíveis para o ndarray.
 - <https://numpy.org/doc/stable/reference/arrays.ndarray.html>

DEMONSTRAÇÃO

- Iremos fazer um processo para demonstrar as principais funcionalidades da biblioteca NumPy, o notebook (caderno de anotação) ficará disponível para download no material de apoio.
- A demonstração será realizada usando o Google Colab.
- Importante que todo processo seja replicado por vocês para que possam aprofundar e solidificar os conhecimentos!
- Demonstração



FINALIZANDO

- Essencial saber explorar o NumPy.
- Não se preocupe muito com os dTypes do NumPy, concentre-se em saber que existem os tipos principais: ponto flutuante, complexo, inteiro, booleano, string e objeto.
- É possível pensar o ndarray com mais de 2 dimensões;
- Repita essa aula quantas vezes for necessário para melhorar a compreensão do funcionamento do NumPy.

INTRODUÇÃO À CIÊNCIA DE DADOS

Biblioteca NumPy