NUMPY, MATPLOTLIB, AND PANDAS

Eric Prebys

Numerical Python (numpy)

- **numpy** is the key the mathematical power of Python
- It allows us to create and manipulate arrays of arbitrary dimension.
- It provides a full mathematical library, parallel to the routines in the math library, which allows the processing of arrays.
- It also provides every possible operation you can think of involving vectors, matrices, etc
- It stores numbers in specific standard formats
  - int16,int32,int64,(int128 on some systems)
  - float16,float32,float64,(float128 on some systems)
- Arrays of numbers are stored in contiguous blocs, compatible with other languages.

Matplotlib

- **matplotlib** is a complex but easy to use set of plotting tools.
- Very simple to make simple plots, but allows full control for more advanced features.
- Includes some basic histogramming tools, but we will soon use the more advanced tools in **scipy**
- Like all packages, works directly with **numpy** arrays.
- Can be augmented with the **seaborn** library for more advanced plots.

Pandas

- **pandas** is a very powerful tool for arranging and manipulating data.
- It works with “data frames”, which can be thought of as internal spreadsheets
- Supports complex mathematical, selection and filtering operations.
- Interfaces directly to plotting packages and other libraries.
- Particularly useful in machine learning

Scientific Python (scipy)

- **scipy** is a diverse library of routines that build on and extend **numpy**
  - Optimization and Fitting (scipy.optimize)
  - Linear Algebra (scipy.linalg)
  - Statistics & Probability (scipy.stats)
  - Integration (scipy.integrate)
  - Signal Processing (scipy.signal)
  - FFT (Fast Fourier Transforms) (scipy.fft)
  - Image Processing (scipy.ndimage)
  - Spatial Data & Geometry (scipy.spatial)
  - Interpolation (scipy.interpolate)
  - Special Functions (scipy.special)
- Too much to talk about all at once. We'll introduce these libraries as needed.