

Python libraries for machine learning

Introduction to Scientific Python

Elham Taghizadeh¹,

¹ College of Convergent Sciences and Technologies, Islamic Azad University Central Tehran Branch

August 30, 2023





Fundemental package for scientific computing with Python

Introduction to Scientific Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises



Introduction to Scientific Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Matplotlib

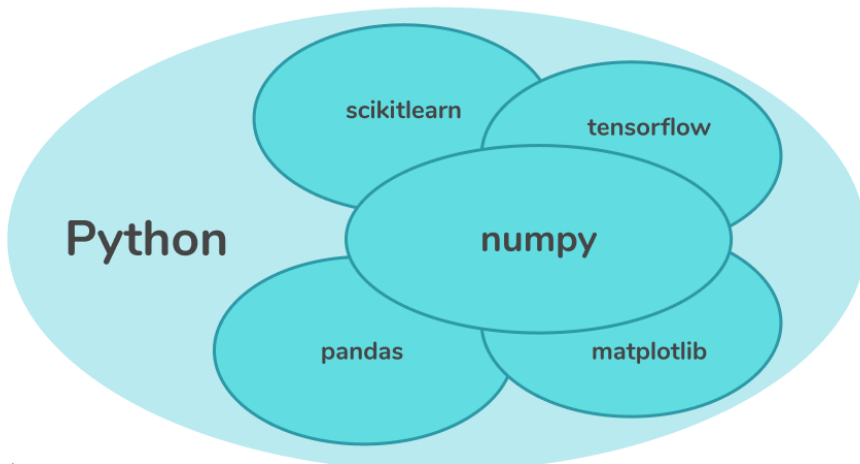
Pandas

Applications

Data analysis

Image Data

Exercises





A number of scientific Python modules have been created and are available in the Python. Some of the most popular modules relevant to this lecture's scope are:

- **numpy:** A powerful library for manipulating arrays and matrices.
- **scipy:** Provides functions for performing higher order mathematical operations such as filtering, statistical analysis, image processing etc.
- **matplotlib:** Provides functions for plotting and other forms of visualization.
- **pandas:** It is used for data manipulation and analysis. It provides special data structures and operations for the manipulation of numerical tables and time series
- **Python Imaging Library:** Provides functions for basic image reading, writing and processing.



Introduction to NumPy

A numpy module adds the ability to manipulate arrays and matrices using a library of high-level mathematical functions. Numpy is derived from the now defunct modules Numeric and Numarray. Numeric was the first attempt to provide the ability to manipulate arrays but it was very slow for computation on large arrays. Numarray, on the other hand, was too slow on small arrays. The code base was combined to create numpy. Numpy has functions and routines to perform linear algebra, random sampling, polynomials, financial functions, set operations etc. Since this lecture is focused on image processing and since images are arrays, we will be using the matrix manipulation capabilities of numpy. The second module that we will be discussing is scipy, which internally uses numpy for its matrix manipulation.

Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

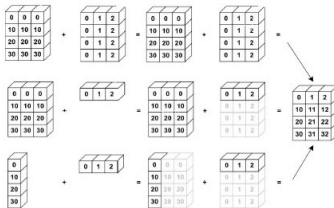
Exercises



- Narray: Narray (or N-d array) is a multidimensional data structure in NumPy. Actually, it is the most important data structure in Scientific Python because all other libraries and data structures in Scientific Python stack use NumPy ndarrays in some form or the other to represent data.
- Installation of NumPy and Matplotlib
- Getting Started with NumPy Programming
- Narray Properties
- Narray Creation Routines
- Basic operations



- Array broadcasting: when operating on two arrays, numpy compares shapes. Two dimensions are compatible when
 - They are of equal size
 - one of them is 1





- Vector operations
 - inner product
 - outer product
 - dot product
- Matrix operations
 - `import numpy.linalg`
 - `eye(3)` identity matrix
 - `trace(A)` Trace
- Linear Algebra
 - `inv(A)` Inverse Matrix
 - `solve(A,b)` Solves $Ax = b$ for A full rank
 - `lstsq(A,b)` Solves $\operatorname{argmin}_x \|Ax - b\|$
 - `eigvals(A)` Computes eigenvalues
 - `pinv(A)` Computes pseudo- inverse of A



- Random Sampling
 - `import numpy.random`
 - `rand(d_0, d_1, \dots, d_n)` Random values in a given shape
 - `randn(d_0, d_1, \dots, d_n)` Random standard normal
 - `randint(l, h, size)` Random integers $[l, h)$
- Distributions in Random
 - `beta`
 - `binomial`
 - `chisquare`
 - `exponential`
 - `dirichlet`
 - `gamma`
 - `laplace`

`file:///C:/Users/West/Downloads/00-NumPy-Arrays(1).html`

Scipy Linear Algebra



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises

Slightly different from `numpy.linalg`.

- Some more functions
- Functions can be slightly different

Scipy Statistics



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises

- Mean, Median, variance
- Gaussian kernel density estimation
- Pearson correlation coefficient
- etc.

Scipy Sparse



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises

- Sparse Matrix classes: CSC, CSR, etc.
- Functions to build sparse matrices
- `sparse.linalg` module for sparse linear algebra
- etc.

Scipy Signal

- Convolutions
- B-Splines
- Filtering
- Wavelets
- Continuous time linear system
- etc.



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises



What is Matplotlib?

We know that matplotlib is the plotting and visualization library of Python Scientific ecosystem. Matplotlib supports NumPy ndarrays and accept them as arguments for its plotting routines.

- Plotting library for Python
- Works well with Numpy
- Syntax similar to Matlab

Let's have a look at few more NumPy ndarray creation routines and matplotlib plotting routines too.

`file:///C:/Users/West/Downloads/00-Matplotlib-Basics.html`



Pandas and why it is important

pandas is a library containing high-level data structures and tools that have been created to assist a Python programmer to perform powerful data manipulations, and discover information in that data in a simple and fast way. The simple and effective data analysis requires the ability to index, retrieve, tidy, reshape, combine, slice, and perform various analyses on both single and multidimensional data, including heterogeneous typed data that is automatically aligned along index labels.

- High performance array and table structures for representation of homogenous and heterogeneous data sets: the Series and DataFrame objects.
- Flexible reshaping of data structure, allowing the ability to insert and delete both rows and columns of tabular data
- Labeling of series and tabular data to facilitate indexing and automatic alignment of data

Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises



- Ability to easily identify and fix missing data
- Extensive I/O facilities to load and save data from multiple formats including CSV, Excel, relational and non-relational databases, and JSON
- Built-in support to retrieve and automatically parse data from various web-based data sources such as Yahoo!, Google Finance, the World Bank, and several others

pandas itself is not a data science toolkit. It does provide some statistical methods as a matter of convenience, but to draw conclusions from data, it leans upon other packages in the Python ecosystem, such as SciPy, NumPy, scikit-learn, and upon graphics libraries such as matplotlib for data visualization. This is actually the strength of pandas over other languages such as R, as pandas applications are able to leverage an extensive network of robust Python frameworks.

`file:///C:/Users/West/Downloads/06-Data-Input-and-Output.html`



Introduction to Scientific Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises

Applications



Math Operations = Data Operations

In the field of data analysis and prediction, operations that produce analyses are typically grounded in mathematical operations, particularly operations in linear algebra. Luckily enough for us, NumPy also supports these functions efficiently. Within the NumPy module, there are tons of matrix operations you can use; and as with any module, this reduces the amount of code you need to write. In this section, we'll review some concrete examples of when these operations occur in the context of a data science-related problem.

Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises



Images are Data

Images consist of pixels, which vary in numerical value. But that's not the important part. The important part is what this structure looks like. Consider this picture of dog:



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises



This image is 200×200 pixels. It's how we described the dimensionality of a matrix. Because of this wonderful property, we can literally treat the pixels of an image as an $n \times n$ matrix.

In the following example, we'll do this using Numpy and Mathplot.

```
file:///C:/Users/West/Downloads/Applications.html
```


Exercises



Introduction
to Scientific
Python

Elham
Taghizadeh

Fundamental
package for
scientific
computing
with Python

Python Modules

Numpy

Scipy

Mathplotlib

Pandas

Applications

Data analysis

Image Data

Exercises

- See course website for exercises for this week
- Get to know the person next to you and do them in pairs
- Let me know if you have any question