

INRAE



université
PARIS-SACLAY

➤ Useful Python packages

Alberto Tonda, Ph.D., Senior permanent researcher (DR)

UMR 518 Mathématiques et Informatique Appliquées - PS, INRAE, U. Paris-Saclay

UAR 3611 Institut des Systèmes Complexes Paris Ile-de-France, CNRS

alberto.tonda@inrae.fr

➤ Another example of title for a slide

- This is some text
 - And some smaller text

➤ Python packages for data science

- One of main strengths of Python is **packages**
 - Lots of good packages for data science
 - Going into the details of each is impossible, read documentation
- This list is from my own personal experience

➤ Basic functionalities

- **datetime** if you need to get information on current time
- **logging** for logging
- **os** for managing paths, creating directories, etc.
- **pickle** for saving complex data structures to disk
- **sys** for stuff related to system

➤ Managing data and plotting

- Managing data and plotting
 - **pandas** for managing CSV/Excel files
 - **matplotlib** and **seaborn** for nice plots
- Basic mathematical functionalities
 - **numpy** and **scipy** for a lot of convenient mathematical functions
 - **sympy** for symbolic computation (e.g. compute derivative)

➤ Machine learning and optimization

- Machine learning
 - **scikit-learn** for machine learning pipelines
 - **PySR** for symbolic regression
 - **xgboost**, **catboost**, and other scikit-learn based packages
 - **pytorch** or **keras** for deep learning
- Optimization
 - **inspyred** for evolutionary optimization
 - **cma** for the CMA-ES
 - **pymoo** for multi-objective optimization

➤ Other stuff

- Regular expressions
 - **re** is the Python module for regular expressions
 - RegExps are complicated to use, excellent for parsing text
 - Describe a pattern to look for in a string, search for pattern
 - I use them often for simple applications

```
# get the description of the state variables, and capture their names
state_variables = regex.findall("([0-9|a-z|\\_]+)\\:\\s+", system["var_description"])
```

- “([0-9|a-z|_]+)\\:\\s+” matches a string containing digits [0-9], lowercase letters [a-z] or underscores [_], repeated any number of times, ending with “:” and any number of spaces after

➤ How to decide if a package is worth using?

- Check popularity and community
 - Simple metric, but adoption usually correlates with usefulness
 - Number of stars on GitHub repo? How many people talk about it?



- Check documentation
 - Is the documentation well done and readable?
 - Are there examples, code that can be cut/pasted?
 - Are authors responsive to issues and pull requests?



INRAE



université
PARIS-SACLAY



alberto.tonda@inrae.fr