

Scientific Python (for those who already know Python basics)

Course Workload		Assessment form (examination/ graded test/ ungraded test)
ECTS	Hours	
3	108	Ungraded test

The course continues students' acquaintance with Python. As part of the course, students will learn how to use various open-source libraries for data analysis, as well as solve more complex bioinformatics problems using Python.

Course structure:

1. Basic text processing with python. Regular expressions

- 1.1. Introduction to re package
- 1.2. Regular expressions. "Greedy" and "Lazy" quantifiers
- 1.3. Regular expressions. Metacharacters, special sequences and sets
- 1.4. Regular expressions. Lookarounds

2. Code organization. Error handling and debugging

- 2.1. Debugging possibilities with python. IDE functions, remote debugging.
- 2.2. Exceptions in python
- 2.3. Best practices to code organization. Philosophy and approached to error handling

3. Exploratory data analysis with python. Data manipulation packages

- 3.1. Table data processing. Pandas package
- 3.2. API requests. The concept of API and common data formats.
- 3.3. Multidimensional array processing. Numpy package
- 3.4. API requests. Python packages and usage examples

4. Visual data analysis with python. Data visualization packages

- 4.1. Interactive plots in python. Plotly package
- 4.2. Core python functionality for visual data analysis. Matplotlib package
- 4.3. Ready-to-use plots for analysis. Seaborn package

4.4. Visual analysis example using python

5. Functional programming with python

- 5.1. Python functional possibilities. Itertools and functools packages
- 5.2. Working with infinite objects. Generators in python
- 5.3. Method oriented programming. Functional programming paradigm
- 5.4. Iterating over objects. Iterators in python

6. Biological data processing with python. Biopython and Scanpy

- 6.1. Biopython package. Working with sequences
- 6.2. Biopython package. Alignment tools
- 6.3. Biopython package. Parsing different data types
- 6.4. Biopython package. Entrez database. Pathway analysis example
- 6.5. Biopython package. Built-in biological data sources.
- 6.6. Single cell analysis example with Scanpy package

7. Scientific project organization. Virtualization, Pipelines and Web applications

- 7.1. Docker. Basic functionality. Running containers
- 7.2. Docker. Using persistence data. Mounting folders
- 7.3. User oriented project design. Host web application using python
- 7.4. OS-level virtualization. Introduction to Docker
- 7.5. Docker. Build your own image
- 7.6. Introduction to pipeline development using Snakemake

8. Machine learning possibilities in python

- 8.1. Linear regression with python. Review of statistical packages
- 8.2. Advanced tools for data processing. Spark and Kafka
- 8.3. Introduction to neural networks with python
- 8.4. Machine learning packages in python. Scikit-learn