Course description

This course is intended for PhD students, engineers and research scientists willing to acquire knowledge in scientific programming. Throughout the course, we will use Python language to lead participants from the basics of computer programming to more advanced techniques such as practical machine learning techniques.

The course is divided into three main topics. We first expose students to the basics of programming using Python language. Then, we discuss about different data formats and data mangling techniques using state-of-the-art scientific packages (e.g., Pandas library). Finally, we will have a tour of machine learning algorithms using the scikit learn package. The exact details of the course will be available online soon (http://c3bi.pasteur.fr/spp).

Syllabus

Chapter 1: Python basics (10 hours)
- General introduction
- Programming (iteration, conditions)
- Python data structures and types
- Functional programming
- Overview of OOP (object oriented programming) in python
Chapter 2 : Data handling (20 hours)
- Introduction to Data Modelling
- Describe a dataset with the help of an Entity-relation model
- Relational database and SQL language (with sqlite)
- Fast and efficient DataFrame in python (with Pandas)

Chapter 3 : Machine learning with scikit learn (20 hours)
- Supervised classification (decision trees, random forest, linear approaches, SVM)
- Regression (Ordinary Least Squares Regression, Elastic Net, Lasso)
- Neural networks and deep learning
- Clustering (k-means, hierarchical clustering)
- Model selection (cross validation, bootstrap, AIC, BIC)