

Genetics and Bioinformatics: an Introduction

Chloé-Agathe Azencott & Thomas Walter

Context

Biology has become a quantitative science. Genomics, in particular, generate large amount of data that cannot be examined “by hand” or “by eye”. These data are heterogeneous in nature (sequences, gene expression, mutation patterns, protein structures, biological images, etc.), and are used to answer very diverse questions.

In this course, you will learn some notions of biology and genomics and discover recent data analyses techniques that can be applied to genetic data. These methods (clustering, classification and regression, image analysis, deep learning, etc.) are generic and can be applied to many other domains.

This course is given (mostly) by researchers from the Centre for Bioinformatics (CBIO) of Mines ParisTech associated with Institut Curie. Most of our examples will therefore be drawn from cancer research applications.

Schedule

All lectures are on Monday afternoons, on 3 time slots: 14:10 – 15:20; 15:30 – 16:40; 16:50 – 18:00

09/16	Notions of biology and genetics – Véronique Stoven
09/23	Notions of biology and genetics – Véronique Stoven
09/30	Machine learning for bioinformatics – Chloé-Agathe Azencott Hands-on session: PSA level prediction – Chloé-Agathe Azencott <i>Bring your own laptop!</i>
10/07	Introduction to microscopy & bioimage informatics – Thomas Walter Computational phenotyping I – Thomas Walter
10/14	Computational phenotyping II – Thomas Walter Hands-on session: deep learning for image recognition – Joseph Boyd
10/21	Gene expression and clustering analyses – Chloé-Agathe Azencott
10/28	Pharmaceutical pipeline – Chloé-Agathe Azencott <i>This session will mostly consist in work in autonomy.</i>
11/04	NIKON Imaging Center visit – Thomas Walter Bioimage informatics – Thomas Walter
12/09	Talk by a bioinformatician – TBD Scientific article discussion – Chloé-Agathe Azencott
12/16	Written exam

Evaluation

- **Hands-on session report: PSA level prediction (15%)**
You will get started on the hands-on session on Sep 30. You will work in Python, on a so-called Jupyter notebook (see below for installation instructions). Your report will take the form of a completed Jupyter notebook and is due on **October 22, 11:59pm**.
- **Pharmaceutical pipeline report (15%)**
This session is an experiment in making your learning process more active. You will be given a set of lecture notes on the pharmaceutical pipeline, and a list of questions and short problems to solve using the lecture notes (and additional resources). You can work together, in class or at home. For the last part of the session (16:50 – 18:00), you will be able to ask questions online to the instructor. The report is due on **October 29, 11:59pm**.
- **Written exam:**
 - **Questions based on the lectures (40%)**
 - **Questions on a scientific article (30%)**
You will be given a scientific article that you will read during the term. We will discuss this article on December 9 – this will be the opportunity for you to ask questions.

Course materials will be made available online at:

https://frama.link/s1133_2019

Scikit-learn installation instructions (before September 30)

The hands-on session on September 30 will be given as a Jupyter notebook, using Python and a few data science / machine learning libraries : numpy, scipy, pandas, matplotlib, and scikit-learn.

If you are familiar with python and comfortable with managing your own installation, make sure you have python3 installed and the following packages (all can be installed with pip: numpy, scipy, pandas, matplotlib, scikit-learn, jupyter and jupyterlab).

If you are not familiar with python and library management, we recommend using either

- miniconda: <https://docs.conda.io/en/latest/miniconda.html>
- anaconda: <https://www.anaconda.com/distribution/>

Miniconda is lighter, but you will need to make sure all the required packages are installed; anaconda is heavier (requires a few GB of space) but everything should work “out of the box”. Please follow the installation instructions for your operating system (Mac/Windows/Linux) and make sure to install the Python3.7 version.

If you're unsure whether your Windows machine is running a 32-bit or 64-bit system, you can use the instructions here: <https://www.lifewire.com/am-i-running-a-32-bit-or-64-bit-version-of-windows-2624475> to check. If you have a 32-bit version, you'll need to use miniconda. For Linux, run “uname -i” in a terminal. If the answer is x86_64, you have a 64-bit system; if it is i386 or i686, you have a 32-bit system.