UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Fundamentals of Bioinformatics and Biostatistics

2425-1-E0201Q081

Aims

The course will introduce the fundamentals of bioinformatics and biostatistics, with a particular focus on biotechnology applications.
The main topics will be related to the computational methods for sequence analysis (with a particular attention to next generation sequencing platforms), genome reconstruction, and protein structure prediction. The course will also introduce the basic concepts of machine learning and artificial intelligence for biological, biotechnological and biomedical applications.
The tools of descriptive and inferential statistics will be explained to guide the student through the typical process at the basis of an experimental study, which starts with the design of the experiment, proceeds with data analysis, and ends with the statistical interpretation and the critical analysis of the relevance of the obtained results.

Knowledge and understanding.
The student will gain knowledge about:
- “computational thinking”, in order to critically use concepts and tools of computer science for the solution of a given problem;
- the choice of the proper sampling and statistical methods, and the interpretation of outcomes after data analysis.

Applying knowledge and understanding.
The student will be able to apply the acquired knowledge for the solution of computational and statistical problems, in addition to:
- development of basic algorithms;
- organizing and handling biological data in automatic ways (practical skills on the use of spreadsheets).

Making judgements.
The student will be able to process the acquired knowledge and choose the proper computational/statistical methods for different applications.

Communication skills.
Use of an appropriate scientific vocabulary and ability in oral/written reports

Learning skills.
Skills in literature reading and understanding, as well as in the elaboration, analysis and application of the acquired knowledge in other courses related to the application of computational and statistical methods for data analysis.

Contents

Bioinformatics
Algorithms, computational thinking, structured programming, computational complexity.
Basics of bio-inspired meta-heuristics, machine learning and artificial intelligence.
Sequence analysis and genome reconstruction (with basics concepts on sequencing platforms).
Computational methods for protein structure prediction.

Biostatistics
Descriptive statistics.
Inferential statistics.

Laboratory
Spreadsheets.
Biological databases.
Application of biostatistics methods.

Detailed program

Bioinformatics

- Computational thinking and basics of structured programming. Definition of algorithm. Structured programming and pseudo-code. Notions of computational complexity.
- Basics of bio-inspired computational methods, supervised and unsupervised machine learning, artificial intelligence.
- Basics of next generation sequencing platforms. Algorithms and heuristics for sequence alignment. Genome reconstruction and annotation. Biological databases.
- Protein structure prediction and molecular interaction (protein folding, molecular docking).

Biostatistics

- Descriptive statistics. Introduction to statistics (types of data, collecting sample data). Summarizing and graphing data (frequency distributions, histograms, boxplots). Statistics for describing, exploring and comparing data (measures of center, measures of variation, measures of relative standing).
- Basic concepts of probability. Probability distributions (uniform, binomial, normal, t distribution). Sampling distributions.
- Inferential statistics. Confidence intervals. Hypothesis testing. Correlation and regression.

Laboratory

- Use of spreadsheets for automatic analysis of different types of data.
- Biological databases (gene, genomes, proteins).
- Application of descriptive and inferential statistics for biological data analysis.

Prerequisites

Background: none.
Prerequisites: none.

Teaching form

Classroom lectures.
20 lectures (2 hours/lectures) composed by:
- a section of delivered didactics (Didattica Erogativa, DE) focused on the presentation-illustration of contents by the lecturer;
- a section of interactive teaching (Didattica Interattiva, DI) including team work and teaching interventions supplementary to delivered didactic activities (case studies, applications of theoretical concepts on real data).
Didactic activities are conveyed by means of face-to-face lectures.

Hands-on sessions in lab room.
10 sessions (3 hours/session) of practical activities in a pc room delivered by interactive teaching through in-person tutorials.

Textbook and teaching resource

All the educational material, including slides and videorecordings, will be available on Moodle platform.

Textbooks:
- M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole. Fondamenti di Bioinformatica. Zanichelli, 2018
- M.M. Triola, M.F. Triola, J. Roy. Fondamenti di statistica per le discipline biomediche (seconda edizione). Pearson, 2022
- M.C. Whitlock, D. Schluter. Analisi statistica dei dati biologici. Zanichelli, 2022

Semester

First semester.

Assessment method

Written exam (90 minutes), consisting in:
- 12 multiple-choice questions (to verify the learning ability acquired by the student), 6 of which related to bioinformatics and 6 to biostatistics;
- 2 open questions (to check the full comprehension and critical analysis capability acquired by the student), 1 of which related to bioinformatics and 1 to biostatistics.
The 14 questions will cover all topics presented during both the classroom lectures and the hands-on sessions.

No mid-term exams will be scheduled.

Office hours

On demand by e-email.

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING | QUALITY EDUCATION | INDUSTRY, INNOVATION AND INFRASTRUCTURE