COURSE SYLLABUS

Fundamentals of Bioinformatics and Biostatistics

2223-1-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 and genome reconstruction (with a particular attention to next generation sequencing platforms), transcriptome analysis and protein structure prediction. The course will also introduce the basic concepts of machine learning and artificial intelligence for biological 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 and ends with the statistical interpretation and the critical analysis of the relevance of the obtained results.

1. 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 in the analysis of biological/clinical data.

2. Applying knowledge and understanding.

The student will be able to apply the knowledge listed in item 1 for the solution of computational and statistical problems, in addition to:
- organizing and handling biological data in automatic ways (practical skills on the use of spreadsheets);
- development of basic algorithms using Python language (hands-on sessions in pc room).

3. Making judgements.

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

4. Communication skills.

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

5. 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.
Next generation sequencing platforms, sequence analysis, genome reconstruction.
Transcriptome analysis.
Protein structure prediction.

Biostatistics:
Descriptive statistics.
Inferential statistics.

Laboratory:
Spreadsheets and biological databases.
Basic notions of programming in Python language.

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.
- Next generation sequencing platforms. Algorithms and heuristics for sequence alignment. Genome reconstruction and annotation. Biological databases.
- Transcriptome analysis, identification of differentially expressed genes.
- 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

Spreadsheets. Biological databases. Basic notions of programming in Python language.

Prerequisites

Background: none.
Prerequisites: none.

Teaching form

Classroom lectures (40 hours) and hands-on sessions in a lab room (30 hours), supported by PowerPoint slides. All lectures and hands-on sessions will be videorecorded and published on the Moodle platform.

Textbook and teaching resource

All the educational material will be available on Moodle platform.

Textbooks:

- P.J. Deitel, H.M. Deitel. Introduzione a Python. Per l'informatica e la data science. Pearson, 2020
- M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole. Fondamenti di Bioinformatica. Zanichelli, 2018
- F. Amaldi, P. Benedetti, G. Pesole, P. Plevani. Tecniche e metodi per la biologia molecolare. Casa Editrice Ambrosiana - Zanichelli, 2020
- 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 (2 hours), consisting in 9 multiple-choice questions and 1 open question about the topics presented during 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