

Course guide

200630 - FBIO - Fundations of Bioinformatics

Last modified: 11/04/2024

Unit in charge: School of Mathematics and Statistics
Teaching unit: 1004 - UB - (ENG)Universitat de Barcelona.

Degree: MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH (Syllabus 2013). (Optional subject).

Academic year: 2024 **ECTS Credits:** 5.0 **Languages:** Spanish

LECTURER

Coordinating lecturer: ESTEBAN VEGAS LOZANO

Others: Primer quadrimestre:
ESTEBAN VEGAS LOZANO - A

REQUIREMENTS

Knowledge of statistical software R.

References:

-R: A self-learn tutorial. <http://www.nceas.ucsb.edu/files/scicomp/Dloads/RProgramming/BestFirstRTutorial.pdf>
-simpleR- Using R for Introductory Statistics: <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

5. CE-1. Ability to design and manage the collection of information and coding, handling, storing and processing it.
6. CE-2. Ability to master the proper terminology in a field that is necessary to apply statistical or operations research models and methods to solve real problems.
7. CE-3. Ability to formulate, analyze and validate models applicable to practical problems. Ability to select the method and / or statistical or operations research technique more appropriate to apply this model to the situation or problem.
8. CE-5. Ability to formulate and solve real problems of decision-making in different application areas being able to choose the statistical method and the optimization algorithm more suitable in every occasion.
Translate to english
9. CE-6. Ability to use appropriate software to perform the necessary calculations in solving a problem.
10. CE-9. Ability to implement statistical and operations research algorithms.

Transversal:

1. ENTREPRENEURSHIP AND INNOVATION: Being aware of and understanding how companies are organised and the principles that govern their activity, and being able to understand employment regulations and the relationships between planning, industrial and commercial strategies, quality and profit.
2. SUSTAINABILITY AND SOCIAL COMMITMENT: Being aware of and understanding the complexity of the economic and social phenomena typical of a welfare society, and being able to relate social welfare to globalisation and sustainability and to use technique, technology, economics and sustainability in a balanced and compatible manner.
3. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.

TEACHING METHODOLOGY

Theory sessions:

In the theory sessions, the professor will present the problems that are tackled in each topic and will provide a summary of the principle concepts and problematic points of each topic.

The student should complete the professor's explanations by consulting the reference texts and complementary materials.

Practical Sessions:

The practical sessions will be conducted with the computer, where instruction will take place regarding the use of bioinformatics tools pertinent to each topic and the problems that are posed.

LEARNING OBJECTIVES OF THE SUBJECT

Upon completing the course, the student must be able to:

- *Identify the bioinformatics domain of study.
- *Know the large group of problems that bioinformatics poses.
- *Be familiar with the most typical methods and models in bioinformatics.

- *Be familiar with the basic components of organisms.
- *Understand the coding and transmission mechanisms of biological information.
- *Know the processes of gene expression and its regulation.

- *Know the existence and availability of diverse information resources, both basic (nucleic acids, proteins, etc.) and more complex (patterns, genomes, etc.).
- *Know the principle tools for recovering information such as Gquery.
- *Know how to access these resources and make queries for obtaining information.

- *Understand and differentiate distinct types of problems related to the alignment of sequences: in pairs, multiples and data search.
- *Know the algorithms for aligning two sequences in optimum form.
- *Know how to perform and interpret the alignment of two sequences.
- *Understand the problem of Multiple Sequence Alignment (MSA).
- *Know how to perform and interpret an MSA.
- *Know how to conduct a sequence search in a database and how to interpret the results.

- *Know the principle methods for representing an MSA and understand the relationships (hierarchical) between them.
- *Understand the basic components of Markov models and their application toward sequence analysis.
- *Know the basic components of a hidden Markov model and understand its advantages and uses for biological problems.

- * Improve R programming skills.
- * Knowing how to create dynamic reports with Rmarkdown.
- * Know some machine learning algorithms for the analysis of biological sequences.

STUDY LOAD

Type	Hours	Percentage
Hours large group	30,0	24.00
Self study	80,0	64.00
Hours small group	15,0	12.00

Total learning time: 125 h

CONTENTS

1. Introduction to Bioinformatics

Description:

Introduction: The explosion of information. About biological Information. But,... What is bioinformatics?. Main thematic of the bioinformatics.

Full-or-part-time: 2h

Theory classes: 2h

2. Basic Concepts of Molecular Biology

Description:

Cells and organisms. Molecules of life (Biomolecules). Central dogma of molecular biology. Genes and gene expression

Full-or-part-time: 5h

Theory classes: 5h

3. Biological Databases: Concepts, Types and Applications

Description:

Bioinformatics and databases. Databases in molecular biology. Format of the stored information.

Full-or-part-time: 9h

Theory classes: 9h

4. Sequence Alignment.

Description:

Introduction to sequence alignment. Types of alignments. Alignments by pairs. BLAST and FASTA. Multiple alignments. Regular expressions and secondary databases

Full-or-part-time: 12h

Theory classes: 12h

5. Probabilistic models of biological sequences.

Description:

Markov chains. Hidden Markov Chains. Applications in bioinformatics

Full-or-part-time: 7h

Theory classes: 7h

6. Machine learning in bioinformatics

Description:

Basic concepts. Algorithms. Application to biological sequences. k nearest neighbors. Support vector machines. Kernels

Full-or-part-time: 10h

Theory classes: 10h

GRADING SYSTEM

The evaluation will be based on four components:

- *Completion of short test exercises (2) during class hours (25%)
- *Class participation and completion of assigned exercises during practice sessions (25%)
- *Presentation of assigned work throughout the course (50%)

BIBLIOGRAPHY

Basic:

- Atwood, T.K.; Parry-Smith, D.J. Introducción a la bioinformática. Madrid: Prentice-Hall, 2002. ISBN 8420535516.
- Claverie, J.M.; Notredame, C. Bioinformatics for dummies [on line]. 2nd ed. New York: Wiley, 2007 [Consultation: 05/07/2023]. Available on : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=284504>. ISBN 0764516965.
- Lee, Jae K. Statistical bioinformatics : for biomedical and life science researchers. Wiley-Blackwell, 2010. ISBN 9780471692720.
- Baldi, Pierre; Brunak, Soren. Bioinformatics : the machine learning approach. 2nd ed. MIT Press, cop. 2001. ISBN 026202506X.

Complementary:

- Durbin, R. [et al.]. Biological sequence analysis : probabilistic models of proteins and nucleic acids [on line]. Cambridge: Cambridge University Press, 1998 [Consultation: 05/07/2023]. Available on : <https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=320915>. ISBN 0521629713.
- Ewens, W. J.; Grant, G. R. Statistical methods in bioinformatics : an introduction. 2nd ed. New York: Springer, 2005. ISBN 0387400826.
- Kohane, I. S.; Kho, Alvin T.; Butte, Atul J. Microarrays for an integrative genomics. Cambridge, Massachusetts: MIT Press, 2003. ISBN 026211271X.
- Mount, David W. Bioinformatics : sequence and genome analysis. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 2004. ISBN 0879696877.
- Gibas, Cynthia; Jambeck, Per. Developing bioinformatics computer skills. Beijing [etc.]: O'Reilly, 2001. ISBN 1565926641.
- Lesk, Arthur M. Introduction to bioinformatics. 3rd ed. Oxford: Oxford University Press, cop. 2008. ISBN 9780199208043.

RESOURCES

Hyperlink:

- Curs d'introducció a la Bioinformàtica. <http://www.ub.edu/stat/docencia/Biologia/introbioinformatica/> - Documents electrònics. Complete Online Bioinformatics Courses/Tutorials
<http://www.med.nyu.edu/rcr/rcr/btr/complete.html> - Enciclopèdies i diccionaris. Bioinformàtica en la Wikipedia
<http://es.wikipedia.org/wiki/Bioinform%C3%A1tica> />
- Llibres Electrònics. Online lectures in Bioinformatics
http://lectures.molgen.mpg.de/online_lectures.html />
The NCBI Bookshelf
<http://www.ncbi.nlm.nih.gov/sites/entrez?db=books> - Organismes i Institucions. The European Bioinformatics Institute
<http://www.ebi.ac.uk/> />
The National Center for Biotechnology Information
<http://www.ncbi.nlm.nih.gov/> />
Instituto Nacional de Bioinformática
<http://www.inab.org/> - Portals temàtics. BIOINFORMATICS.CA
<http://bioinformatics.ca/> />
123Genomics
<http://www.123genomics.com/> - Revistes. Bioinformatics
<http://bioinformatics.oxfordjournals.org/> />
Briefings in Bioinformatics
<http://bib.oxfordjournals.org/> />
BMC Bioinformatics
<http://www.biomedcentral.com/bmcbioinformatics/> - Webs. Internationals Society for Computational Biology (ISCB)
<http://www.iscb.org/> />
The Gene Discovery Page
<http://www.biowriters.com/bioinformatics/gdp.html>

Other resources:

Bioinformatics notes, available on the intranet or supplied by the professor in pdf.