Course guide
270965 - BSG - Bioinformatics and Statistical Genetics

Unit in charge: Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science.
715 - EIO - Department of Statistics and Operations Research.

Degree: MASTER'S DEGREE IN DATA SCIENCE (Syllabus 2021). (Optional subject).
Academic year: 2022  ECTS Credits: 6.0  Languages: English

LECTURER

Coordinating lecturer: GABRIEL ALEJANDRO VALIENTE FERUGLIO
Others: Primer quadrimestre:
JAN GRAFFELMAN - 10
GABRIEL ALEJANDRO VALIENTE FERUGLIO - 10

PRIOR SKILLS

Basic knowledge of algorithms and data structures.
Basic knowledge of statistics.
Basic knowledge of the Python programming language.
Basic knowledge of the R programming language.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CE2. Apply the fundamentals of data management and processing to a data science problem
CE5. Model, design, and implement complex data systems, including data visualization
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science

Generical:
CG4. Design and implement data science projects in specific domains and in an innovative way

Transversal:
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Basic:
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
TEACHING METHODOLOGY

All classes consist of a theoretical session (a lecture in which the professor introduces new concepts or techniques and detailed examples illustrating them) followed by a practical session (in which the students work on the examples and exercises proposed in the lecture). On the average, two hours a week are dedicated to theory and one hour a week to practice, and the professor allocates them according to the subject matter. Students are required to take an active part in class and to submit the exercises at the end of each class.

LEARNING OBJECTIVES OF THE SUBJECT

1. Introduce the student to the algorithmic, computational, and statistical problems that arise in the analysis of biological data.
2. Reinforce the knowledge of discrete structures, algorithmic techniques, and statistical techniques that the student may have from previous courses.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours large group</td>
<td>54,0</td>
<td>36.00</td>
</tr>
<tr>
<td>Self study</td>
<td>96,0</td>
<td>64.00</td>
</tr>
</tbody>
</table>

Total learning time: 150 h

CONTENTS

**Introduction to bioinformatics**

Description:

**Phylogenetic reconstruction I**

Description:

**Agreement of phylogenetic trees**

Description:

**Phylogenetic reconstruction II**

Description:
Agreement of phylogenetic networks

Description:

Phylogenetic reconstruction III

Description:

Phylogenetic and taxonomic reconstruction

Description:
Phylogenies and taxonomies. Classification of metagenomic samples. Agreement of classifications.

Introduction to statistical genetics

Description:

Hardy-Weinberg equilibrium

Description:

Linkage disequilibrium

Description:

Phase estimation

Description:

Population substructure

Description:
Genetic association analysis

Description:

Family relationships and allele sharing

Description:
Identity by state (IBS) and Identity by descent (IBD). Kinship coefficients. Allele sharing. Detection of family relationships. Graphical representations.

ACTIVITIES

Development of syllabus topics

Specific objectives:
1, 2

Related competencies:
CG4. Design and implement data science projects in specific domains and in an innovative way
CE1. Develop efficient algorithms based on the knowledge and understanding of the computational complexity theory and considering the main data structures within the scope of data science
CE6. Design the Data Science process and apply scientific methodologies to obtain conclusions about populations and make decisions accordingly, from both structured and unstructured data and potentially stored in heterogeneous formats.
CE5. Model, design, and implement complex data systems, including data visualization
CE9. Apply appropriate methods for the analysis of non-traditional data formats, such as processes and graphs, within the scope of data science
CE2. Apply the fundamentals of data management and processing to a data science problem
CT4. INFORMATION LITERACY: Capacity for managing the acquisition, the structuring, analysis and visualization of data and information in the field of specialisation, and for critically assessing the results of this management.
CT5. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB10. Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 105h
Theory classes: 15h
Laboratory classes: 30h
Self study: 60h

Final exam

Full-or-part-time: 33h
Theory classes: 3h
Self study: 30h
GRADING SYSTEM

Students are evaluated during class, and in a final exam. Every student is required to submit one exercise each week, graded from 0 to 10, and the final grade consists of 50% for the exercises and 50% for the final exam, also graded from 0 to 10.

BIBLIOGRAPHY

Basic:

Complementary:

RESOURCES

Hyperlink:
- http://rosalind.info/
- http://www.r-project.org/