270656 - BSG - Bioinformatics and Statistical Genetics

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 715 - EIO - Department of Statistics and Operations Research
723 - CS - Department of Computer Science

Academic year: 2018
Degree: MASTER'S DEGREE IN INNOVATION AND RESEARCH IN INFORMATICS (Syllabus 2012). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan

Prior skills

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

Degree competences to which the subject contributes

Basic:
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.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

Specific:
CEC1. Ability to apply scientific methodologies in the study and analysis of phenomena and systems in any field of Information Technology as well as in the conception, design and implementation of innovative and original computing solutions.
CEC2. Capacity for mathematical modelling, calculation and experimental design in engineering technology centres and business, particularly in research and innovation in all areas of Computer Science.
CEC3. Ability to apply innovative solutions and make progress in the knowledge that exploit the new paradigms of Informatics, particularly in distributed environments.

Generical:
CG3. Capacity for mathematical modeling, calculation and experimental designing in technology and companies engineering centers, particularly in research and innovation in all areas of Computer Science.

Transversal:
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

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>Total learning time: 150h</th>
<th>Theory classes: 24h</th>
<th>16.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practical classes: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Laboratory classes: 12h</td>
<td>8.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 6h</td>
<td>4.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 96h</td>
<td>64.00%</td>
</tr>
</tbody>
</table>
## Content

### Introduction to Bioinformatics

**Degree competences to which the content contributes:**

**Description:**

### Linear programming and integer linear programming

**Degree competences to which the content contributes:**

**Description:**
Solving easy and hard feasibility and optimization problems in Bioinformatics. Linear programming and integer linear programming in R.

### The longest common substring problem

**Degree competences to which the content contributes:**

**Description:**

### The shortest common superstring problem

**Degree competences to which the content contributes:**

**Description:**
The shortest common superstring problem: Assembling short DNA sequence reads. Integer programming formulations.

### The closest and the farthest string problems

**Degree competences to which the content contributes:**

**Description:**
The closest and the farthest string problems: Finding patterns that occur, or do not occur, in each string in a given set of DNA sequences. Integer programming formulations.

### The closest and the farthest substring problems

**Degree competences to which the content contributes:**
**Other string selection problems**

**Degree competences to which the content contributes:**

**Description:**
Other string selection problems: The Close to Most and the Far from Most Strings problems. Integer programming formulations.

---

**Introduction to statistical genetics**

**Degree competences to which the content contributes:**

**Description:**

---

**Hardy-Weinberg equilibrium**

**Degree competences to which the content contributes:**

**Description:**

---

**Linkage disequilibrium**

**Degree competences to which the content contributes:**

**Description:**

---

**Phase estimation**

**Degree competences to which the content contributes:**

**Description:**
Population substructure

Degree competences to which the content contributes:

Description:

Genetic association analysis

Degree competences to which the content contributes:

Description:

Family relationships and allele sharing

Degree competences to which the content contributes:

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

Planning of activities

<table>
<thead>
<tr>
<th>Development of syllabus topics</th>
<th>Hours: 56h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 42h</td>
<td></td>
</tr>
<tr>
<td>Practical classes: 0h</td>
<td></td>
</tr>
<tr>
<td>Laboratory classes: 0h</td>
<td></td>
</tr>
<tr>
<td>Guided activities: 0h</td>
<td></td>
</tr>
<tr>
<td>Self study: 14h</td>
<td></td>
</tr>
</tbody>
</table>

Specific objectives: 1, 2

<table>
<thead>
<tr>
<th>Final exam</th>
<th>Hours: 31h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 3h</td>
<td></td>
</tr>
<tr>
<td>Practical classes: 0h</td>
<td></td>
</tr>
<tr>
<td>Laboratory classes: 0h</td>
<td></td>
</tr>
<tr>
<td>Guided activities: 0h</td>
<td></td>
</tr>
<tr>
<td>Self study: 28h</td>
<td></td>
</tr>
</tbody>
</table>
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:

Others resources:

Hyperlink
- http://www.r-project.org/
- https://ampl.com/