

Guia docent

270656 - BSG - Bioinformàtica i Genètica Estadística

Última modificació: 20/07/2020

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.
723 - CS - Departament de Ciències de la Computació.

Titulació: MÀSTER UNIVERSITARI EN INNOVACIÓ I RECERCA EN INFORMÀTICA (Pla 2012). (Assignatura optativa).

Curs: 2020 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: GABRIEL ALEJANDRO VALIENTE FERUGLIO

Altres: Primer quadrimestre:
JAN GRAFFELMAN - 10
GABRIEL ALEJANDRO VALIENTE FERUGLIO - 10

CAPACITATS PRÈVIES

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

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CEC1. Capacitat per aplicar el mètode científic en l'estudi i anàlisi de fenòmens i sistemes en qualsevol àmbit de la Informàtica, així com en la concepció, disseny i implantació de solucions informàtiques innovadores i originals.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CEC3. Capacitat per aplicar solucions innovadores i realitzar avanços en el coneixement que explotin els nous paradigmes de la Informàtica, particularment en entorns distribuïts.

Genèriques:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Transversals:

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Bàsiques:

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB9. Que els estudiants posseïxin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdrida o autònoma.



METODOLOGIES DOCENTS

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.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

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.

HORES TOTALS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Grup mitjà/Pràctiques	12,0	8.00
Activitats dirigides	6,0	4.00
Aprenentatge autònom	96,0	64.00
Grup gran/Teoria	24,0	16.00
Grup petit/Laboratori	12,0	8.00

Dedicació total: 150 h

CONTINGUTS

Introduction to Bioinformatics

Descripció:

Introduction to the R language. Computational linear algebra. Numerical optimization. Bioinformatics with R.

Linear programming and integer linear programming

Descripció:

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

The longest common substring problem

Descripció:

The longest common substring problem: Finding consensus among DNA sequences. Integer programming formulations.

The shortest common superstring problem

Descripció:

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



The closest and the farthest string problems

Descripció:

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

Descripció:

The closest and the farthest substring problems: Finding short strings that are, or are not, enriched in each string in a given set of DNA sequences. Integer programming formulations.

Other string selection problems

Descripció:

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

Introduction to statistical genetics

Descripció:

Basic genetic terminology. Population-based and family-based studies. Traits, markers and polymorphisms. Single nucleotide polymorphisms and microsatellites. R-package genetics.

Hardy-Weinberg equilibrium

Descripció:

Hardy-Weinberg law. Hardy-Weinberg assumptions. Multiple alleles. Statistical tests for Hardy-Weinberg equilibrium: chi-square, exact and likelihood-ratio tests. Graphical representations. Disequilibrium coefficients: the inbreeding coefficient, Weir's D. R-package HardyWeinberg.

Linkage disequilibrium

Descripció:

Definition of linkage disequilibrium (LD). Measures for LD. Estimation of LD by maximum likelihood. Haplotypes. The HapMap project. Graphics for LD. The LD heatmap.

Phase estimation

Descripció:

Phase ambiguity for double heterozygotes. Phase estimation with the EM algorithm. Estimation of haplotype frequencies. R-package haplo.stats.



Population substructure

Descripció:

Definition of population substructure. Population substructure and Hardy-Weinberg equilibrium. Population substructure and LD. Statistical methods for detecting substructure. Multidimensional scaling. Metric and non-metric multidimensional scaling. Euclidean distance matrices. Stress. Graphical representations.

Genetic association analysis

Descripció:

Disease-marker association studies. Genetic models: dominant, co-dominant and recessive models. Testing models with chi-square tests. The alleles test and the Cochran-Armitage trend test. Genome-wide association tests.

Family relationships and allele sharing

Descripció:

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

ACTIVITATS

Development of syllabus topics

Objectius específics:

1, 2

Competències relacionades:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CEC3. Capacitat per aplicar solucions innovadores i realitzar avanços en el coneixement que explotin els nous paradigmes de la Informàtica, particularment en entorns distribuïts.

CEC1. Capacitat per aplicar el mètode científic en l'estudi i anàlisi de fenòmens i sistemes en qualsevol àmbit de la Informàtica, així com en la concepció, disseny i implantació de solucions informàtiques innovadores i originals.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi.

Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autòdridigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

Dedicació: 56h

Grup gran/Teoria: 42h

Aprenentatge autònom: 14h



Final exam

Dedicació: 31h

Grup gran/Teoria: 3h

Aprenentatge autònom: 28h

SISTEMA DE QUALIFICACIÓ

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.

BIBLIOGRAFIA

Bàsica:

- Pappalardo, E.; Pardalos, P.M.; Stracquadanio, G. Optimization approaches for solving string selection problems [en línia]. Springer, 2013 Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=1538891>. ISBN 9781461490531.
- Braun, W.J.; Murdoch, D.J. A first course in statistical programming with R. 2nd ed. Cambridge University Press, 2016. ISBN 9781107576469.
- Foulkes, A.S. Applied statistical genetics with R: for population-based association studies. Springer Verlag, 2009. ISBN 9780387895536.
- Gusfield, D. Integer linear programming in computational and systems biology: an entry-level text and course. Cambridge University Press, 2019. ISBN 9781108421768.

Complementària:

- Paradis, E. Analysis of phylogenetics and evolution with R [en línia]. Springer, 2012 Disponible a: <https://doi.org/10.1007/978-1-4614-1743-9>. ISBN 9781461417439.
- Gentleman, R. R programming for bioinformatics. CRC Press, 2009. ISBN 9781420063677.
- Weir, B.S. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, 1996. ISBN 0878939024.
- Laird, N.M.; Lange, C. The fundamentals of modern statistical genetics [en línia]. Springer, 2011 Disponible a: <https://doi.org/10.1007/978-1-4419-7338-2>. ISBN 9781441973375.

RECURSOS

Enllaç web:

- <http://www.r-project.org/>. The R Project for Statistical Computing
- <https://ampl.com/>. AMPL