Introduction to Computational Biology for Physicists
01:750:431
Fall 2020

Instructor: Gyan Bhanot - gyanbhanot@gmail.com (preferred) or gbhanot@physics.rutgers.edu

Summary: After a general introduction to the relevant ideas in biology, students will learn the mathematical methods used to analyze genetic and genomic data, including but not limited to, Probability Theory, the Theory of Distributions and Moments, The Central Limit theorem, Linear and non-Linear Regression, Parametric and Non-Parametric Tests of Significance and Analysis of Variance (ANOVA), Sequence Alignment, Phylogenetic Analysis, Clustering and Pattern Recognition, Neural Networks, Monte Carlo Simulations & Evolutionary Game Theory. Given the ongoing Covid-19 pandemic, we will discuss numerical methods to analyze the available data on infections and death and develop a model to understand the dynamics of this pandemic. All the methods and ideas presented will be developed using concrete examples of how they apply to biological phenomena.

The Programming Language Matlab will be used as a computational tool. Students will be taught how to solve many problems by writing Matlab code.

Pre-requisites: Calc2 (01:640:136 or 01:640:138 or 01:640:152)

Co-requisites: None

Meeting times: Two 80 minute lectures per week.
Lecture: M-Th 2 (10:20 am - 11:40 am)
Recitations: None
Online Office Hours: By arrangement

Text: Lecture notes which cover all course content will be provided a week in advance of each class. The textbook for the statistical analysis portion of the course will be "Mathematical Statistics and Data Analysis by John A Rice". A pdf copy of this book will be made available to students before the first day of classes. Some topics will have assigned reading of relevant literature which will be provided as pdf files.

LMS: Canvas

Provisional Plans for Remote Instruction:

- Detailed notes for each lecture will be provided to all students one week in advance. The students will be expected to have read these notes before class.
- Each lecture will start with a 20 minute summary of the lecture notes (may be a taped lecture) followed by 10 minutes to address student questions.
- For the remainder of the class the students will work on a worksheet covering the material. The instructor will be available to answer questions during this time. These

worksheets must be e-mailed to the instructor at the end of the class. The worksheets will count for 20% of the grade.

- Home-work on the material covered during the week will be posted on Friday and will be due one week later. The students will be divided into dynamic work-groups of 4-5 students, which will change each week. Each group will submit one solution for each homework. They will also grade the other members of the group for the degree of participation in the working group. The homework will count as 30% of the grade.
- In addition to the classroom instruction there will be 2 x 2 hour periods when the instructor will have online office hours where students can call in and ask questions.
- Additional individual contact times to address other student concerns will be by arrangement.
- Students will be required to write a term paper on a topic they will choose from a list provided by the instructor after the class has been in session for 6 weeks. This will count as 10% of the grade.
- There will be one mid-term and one final – both will be multiple-choice with a strict time requirement. Each will be 20% of the grade.

Summary of grading:

- In class worksheets : 20% of grade
- Homework: 30% of grade
- Term paper: 10% of grade
- Midterm: 20% of grade
- Final: 20% of grade

Schedule (provisional): Lecture Topics

Week 1-4

- Introduction to Probability and Bayes Theorem, Random Variables; Expected Value and Variance
- Distribution Theory - Binomial, Poisson, Bernoulli & Geometric Distributions
- Matlab Tutorial - Demonstration of Central Limit Theorem.
- Parametric Tests of Significance based on the Central Limit Theorem (t-test, F-test, ANOVA)
- Non-parametric tests of significance

Week 5-9

- Bio Intro – the Genetic Code, Mutation and Drift, Hardy Weinberg Theory
- The role of recombination and selection in the evolution of organisms.
- Introduction to Viruses - FLU, HIV, SARS, MERS and Zoonotic diseases
- Analytical Modeling – The SIR Model of Pandemics - Modeling Covid-19 data
- Monte Carlo Simulations
- Mid-term and assignment of term paper topics

Week 10-14

- Sequence Alignment and Phylogenetics
- Clustering Methods: k-means clustering, PCA, t-SNE and non-negative matrix factorization methods.
- Analysis of Genetic and Genomic cancer data using the techniques learned.

- Neural Networks
- Evolutionary Game Theory
- Collection of Term Papers and Final

Academic Integrity:

Students are expected to maintain the highest level of academic integrity. You should be familiar with the university policy on academic integrity: http://academicintegrity.rutgers.edu/academic-integrity-policy/ Violations will be reported and enforced according to this policy.

Use of external sources to obtain solutions to homework assignments or exams is cheating and a violation of the University Academic Integrity policy. Cheating in the course may result in penalties ranging from a zero on an assignment to an F for the course, or expulsion from the University. Posting of homework assignments, exams, recorded lectures, or other lecture materials to external sites without the permission of the instructor is a violation of copyright and constitutes a facilitation of dishonesty, which may result in the same penalties as explicit cheating.

Not only does the use of such sites violate the University's policy on Academic Integrity, using such sites interferes with your achievement of the learning you are paying tuition for. Assignments, quizzes, and exams are given not simply to assign grades, but to promote the active learning that occurs through completing assignments on your own. Getting the right answer is much less important than learning how to get the right answer. This learning is critical to your success in subsequent courses and your careers.

Student wellness Services
Student Counseling, ADAP & Psychiatric Services (CAPS) wellness for non-emergency psychological health issues services (848) 932-7884, 17 Senior Street, New Brunswick, NJ 08901 http://health.rutgers.edu/medical-counseling-services/counseling/

Violence Prevention & Victim Assistance (VPVA), (848) 932-1181, 3 Bartlett Street, New Brunswick, NJ 08901, http://www.vpva.rutgers.edu/

Office of Disability Services (848) 445-6800, Lucy Stone Hall, Suite A145, Livingston, 54 Joyce Kilmer Avenue, Piscataway, NJ 08854, https://ods.rutgers.edu/

Scarlet Listeners for confidential peer counseling and referral hotline, (732) 247-5555, http://www.scarletlisteners.com