Syllabus: CmpSci 4370

Introduction to Biological Data Science

Department of Mathematics and Computer Science

Semester

SP 2019

Instructor Details

Name: *Dr. Sharlee Climer*
Office Hours: *Mondays and Wednesdays 5:15 – 6:45 or by appointment*
Office Location: *319 ESH*

Submission and Communication

Preferred communication is via email: climer@umsl.edu. Submissions will generally be made on Canvas; procedures and late submission policies will be indicated for each assignment.

Scoring

See table below for scoring scheme.

Incremental Grading

Grades within 2% of the cutoff will be assigned '+' or '-' accordingly, with the exception of grades above 98%, as the university does not allow 'A+' grades.

Schedule

Exams will be held on February 27th and April 10th. A comprehensive final exam will be from 2:45 to 4:45 on Many 13th. All exams will be in our usual classroom. If you have a special circumstance and are unable to take a test on the specified date and time, it must be individually discussed with the instructor and prearranged before the test. ***Documentation will be required.*** **Failure to appear for an exam will result in a grade zero unless prior arrangements have been made.** See below for tentative schedule of class topics.

Course Details

General Policies

Lecture recordings, audio or video, are not permitted unless the instructor explicitly allows that.

We follow the university policies regarding excused EX and EX-F drops.

Students are given and are expected to sustain positive learning environment in class. This means positive conduct in class, no late walk-ins or early walk outs without a good explanation or a prior arrangement, and if on-line access is

available in class - not using it for anything not class related. Cell phones must be turned off during class. Students not meeting these standards may be asked to leave the classroom.

All in and out of class work for grade should be done independently (except for group projects). Homework can be discussed with others, but the final work (code, answer, etc.) should be independent. Programs may be discussed up to design, but no code is allowed to be shared except for what is presented in class. Help can always be sought and received. However, help to assignments should be generic on the subject matter or very narrowly focused on specific problem not being the central point in the assignment.

Course Description

Prerequisites: CMP SCI 3130 or consent of instructor. This course provides an introduction into several key areas of biological data science, with a focus upon genetic data. Relevant background topics in genetics, current issues, and a variety of available resources will be explored. Upon successful completion of this course, the student will be able to evaluate algorithms for analyzing genetic data, including assessments of sources of errors and analysis of time and space complexity; address shortcomings in existing approaches; and implement efficient and effective software for exposing information hidden in genetic data. Credit cannot be granted for both CMP SCI 4370 and CMP SCI 5370.

Text and Other Materials

No textbook is required. Course readings and videos will be provided.

Course Schedule

Schedule is tentative and subject to change.

WEEK	TOPIC
1	Introduction, network modeling
2	Breadth-first search, *k*-means for weighted and unweighted edges
3	Hierarchical clustering, principal component analysis, modularity
4	Spectral clustering, genetic algorithms
5	Mixed-integer programming
6	Review and **Exam 1**
7	Student presentations of project proposals
8	Data pre-processing, Mendelian genetics, Punnett squares
9	DNA, mutations, DNA translation, protein folding
10	Mitosis and meiosis, recombination, linkage disequilibrium
11	Review and **Exam 2**
12	Combinatorial GWAS, Gene co-expression analysis
13	Protein structure analysis, research trends
14	Research trends
15	Student presentations of projects
16	**Final Exam**

Course Objectives and Learning Outcome

Upon successful completion of this course, students should be able to:

- Understand relevant biological terminology and processes
- Understand a number of computational problems faced by biologists
- Utilize existing software for analyzing data
- Evaluate algorithms, including assessments of sources of errors and analysis of time and space complexity
- Identify and address shortcomings in existing approaches
- Implement efficient and effective software for extracting information from biological data
- Understand privacy concerns and ethical responsibilities when handling biological data

Course Grading

We will use the standard 10% grading scale: 90% and above for an A, 80% and above for a B, 70% and above for a C, 60% and above for a D, else F. All grades throughout the course will be posted on Canvas.

Homeworks, quizzes, etc.	10%
Projects	30%
Research trends activities	10%
Two exams (15% each)	30%
Comprehensive final	20%

University Policies and Information

http://www.umsl.edu/~webdev/mathematics/files/pdfs/cs_umsl_syllabus_university.pdf