San José State University
Computer Science Department
CS223 Bioinformatics, Sec 01, Fall 2024

Course Information

Instructor:	Leonard Wesley
Office Location:	MH 212
Telephone:	408.924.5287 (Office, however, I will not be on campus much during the Fall 2024 semester.)
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Tuesdays 7:00AM – 9:00AM, Zoom Link For Office Hours For Fall 2024: https://sjsu.zoom.us/j/82185188270?pwd=UYdaagroNO8GC9lhQBlW1sZh7YzRPO.1 PASSCODE: 279456
Class Days/Time:	Section 01: Tuesdays and Thursdays 3:00PM – 4:15AM
Classroom:	ISB 876
Prerequisites:	CS/BIOL/SE 123B. Allowed Declared Major: Computer Science, Bioinformatics, Data Science.

Official SJSU Catalogue Course Description

The course investigates algorithms for solving computational problems in bioinformatics, and the life processes underlying those algorithms. Metagenomics, bioinformatic machine learning, single-cell RNA sequencing, variant discovery. Possible additional topics: genome assembly, advanced phylogenetics, long-read sequencing, and structural bioinformatics.

Expanded Course Description

The course presents some of the methods and techniques of algorithm analysis that can be used to assess time and space complexity of several algorithms that are used

in bioinformatics. Developed time and space complexity analysis skills will be used to design and assess the performance of epigenetic analysis algorithms that are sued in an evolving pipeline that is used to help identify epigenetic cause of a disease. Data for epigenetic analysis will come, in part, from the results of single cell sequencing. Predictions will be made, in part, by evidential reasoning extensions to HMMs in order to achieve increased fidelity of predictions.

Learning Outcomes

Upon successful completion of this course, students will:

1. SLO-1 DEVELOPING BIOINFORMATICS COMPUTATIONAL SKILLS: Know how to analyze and predict the performance of bioinformatic algorithms, e.g., analyze recursive and iterative implementations of various algorithms used for correlation analysis, PCA, pattern analysis, …, and how they can be used to help answer bioinformatics questions of interest.

2. SLO-2 EVIDENTIAL REASONING AND BIOINFORMATICS: Introduction to the evidential reasoning (ER) calculus that is a generalization of traditional probabilistic and statistical inference. Examples of how ER can be used to help answer bioinformatics related questions with greater fidelity than traditional methods will be presented.

3. SLO-3 ADVANCED HMMs: Discuss some of the limitations of traditional HMMs. Introduce advanced HMMs such as profile HMMs, jumping HMMs, pair-HMMs, sub-HMMs, and phylo-HMMs.

4. SLO-4 EPIGENETICS: Have a sufficient introductory understanding of epigenetics in order to equip students with the knowledge needed to investigate diseases that are caused by epigenetic mechanisms and being able to development a simple analysis pipeline that will be used throughout the remaining semester and integrated with the remaining course topics.

Each SLO above corresponds to a learning module that is described in the course calendar below. That is, there are four (4) learning modules that corresponds to each of the SLOs described above.

Required Texts/Readings

All required text, publications, reference material, and so forth will be provided to the class.

Other Optional Reading Material

Developing Bioinformatics Computer Skills, Cynthia Gibas and Per Jambeck, O'Reilly & associates. (A good book for beginners)

Introduction to Computational Biology: Maps, Sequences and Genomes, Michael S. Waterman, CRC Press. (A statistical oriented view of bioinformatics)

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Andreas D. Baxecvanis and B.F. Francis Ouellette, John Wiley & Sons 2nd Ed. (Includes contributions from several authors providing a wide perspective)

Computational Resources:

Students are required to make sure that they have access to sufficient UNIX, Windows, or Mac based computational resources (e.g., computers and software) to carryout assignments in the course. Likely programming tools used during the course will be Python version 3.9 or later and R version 4.3.2 or later and R-Studio version 2024.2 Build 764 or later. An attempt to offer the course in a classroom with sufficient computation resources will be made by the department to support classroom instruction and demonstrations. However, students should be prepared to bring their portable laptops to class.

Course Requirements and Assignments

Course Logistics

Students should expect to spend approximately nine (9) hours per week (on average) outside of the classroom preparing for and completing the assigned course work. This includes reading papers, viewing videos as appropriate, completing homework and programming exercises, and so forth. The amount of time that a student actually spends studying and completing course work will depend on individual skills and the time that the student actually allocates to the course. The nine (9) hours per week estimate is based on previous experiences of the instructor and students. So please plan and schedule accordingly.

Previously, some students have asked for special exceptions to policies and procedures for this course. An example includes asking the instructor for extra assignments or work to help improve a grade. Even if such a request is reasonable in the opinion of the instructor, no exception will be given to a student unless the same opportunity can be made available to the entire class, and does not constitute significant extra work on the part of students, instructors, graders and so forth. Students should have no concern that other students will receive special exceptions that will not be made available to the entire class.

NOTE: University policy (F69-24) states that "Students should attend all meetings of their classes, not only because they are responsible for material discussed therein, but because active participation is frequently essential to insure maximum benefit for all members of the class. Attendance per se shall not be used as a criterion for grading." However, attendance will be required in order to complete and submit many in-class exercises, quizzes, and exams. Should students miss or leave early from one or more classes, students are responsible for knowing and understanding any and all course subject matter, assignments, exercises, instructions and so forth that are presented or discussed during official scheduled class time.

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the Syllabus/Syllabus.

Quizzes and Exams

There will be three quizzes, one midterm and three "topic-projects" that will replace a final exam all of which will count toward a student's final grade as specified in the "Grades" section below. During quizzes and exams, communication with other individuals via any means is strictly prohibited without the express permission of the instructor. Violations will be met with the full impact of SJSU's academic integrity policy and procedures.

Projects

Several life-science related project topics will be described near the start of the course. Projects will involve applying the skills and knowledge learned in the course to the project. Projects in this course will be individual (not team) projects. Project scores will count toward the final grade as specified in the "Grades" section below.

In-Class Exercises

There will be four in-class exercises where groups of two to four will be formed to work on an assigned exercise. In-class participation is **mandatory**, and an attendance sign-up sheet will be passed around to verify participation. The assigned exercises are intended to reinforce learning and understanding of previous lecture, homework, and programming assignment subject matter by providing hands-on experience with completing the provided assignment. A supplement document named "In-Class Exercise Procedure.pdf" is available on Canvas that describes the general organization and procedure all students are to follow for all in-class exercise assignments. The "In-Class Exercise Procedure.pdf" document should be treated as part of the Syllabus for this course.

Reading, Homework, Programming, Participation Assignments

Graded reading, homework, programming, class participation and brief course feedback assignments will be given almost weekly. For homework assignments, only one or two questions will be graded. However, answers for all questions will be provided no sooner

than four (4) days after the due date. For non-programmer types (e.g., Biology. Biochem, … majors), comparable non-programming tasks will be assigned for programming assignments. All graded assignments will count toward a student's final course grade. Programmer types (e.g., CS, Bioinformatics, Software Engineering,… majors) must submit programming assignments as per specifications that are described in the "Programming Assignment Guidelines.pdf" document on Canvas.

Computational Resources

Students are required to make sure that they have access to sufficient UNIX, Windows, or Mac based computational resources (e.g., computers and software) to carryout assignments in the course. An attempt to offer the course in a classroom with sufficient computation resources will be made by the department to support classroom instruction and demonstrations. However, students should be prepared to bring their portable laptops to class.

Questions and Regrade Requests

All questions about grading and re-grade requests must be presented to the instructor within two weeks from the date the graded assignments, exercises, and exams are returned to the class or by the last day of instruction for the semester (whichever is sooner). Assignments, quizzes, and exams will typically be returned (i.e., posted) to Canvas, or manually handed back in class. General questions about the topics covered in assignments, exams, exercises, programming assignments, and the course are permissible at any time.

Tentative course calendar of assignment due dates & exam dates:

(Please note that course calendar below, and its content is "subject to change with fair notice")

Week and Class Mtg #	Tue	Thur	Module # & Name	TOPIC	Assignment See Canvas For Module & Weekly Assignment Details and Due Dates

Week 1		8/22	#1 DEVELOPING BIOINFORMATICS COMPUTATIONAL SKILLS	8/22: Intro To Course: -Topics, learning objectives, course logistics, Instructor background - Syllabus	Learning Module #1
Week 2	8/27	8/29	#1 DEVELOPING BIOINFORMATICS COMPUTATIONAL SKILLS	8/27: - Analysis of algorithms 8/29: - Time & order complexity using sorting algorithms as examples.	Learning Module #1
Week 3	9/3	9/5	#1 DEVELOPING BIOINFORMATICS COMPUTATIONAL SKILLS	9/3: - Order of Growth vs Asymptotic Growth 9/5: - Master Theorem	Learning Module #1
Week 4	9/10	9/12	#1 DEVELOPING BIOINFORMATICS COMPUTATIONAL SKILLS	9/10: - Master Theorem (cont.) 9/24: - **In-Class Exercise 1 Covers topics From Week1 thru Week 4**	Learning Module #1

Week 5	9/17	9/19	#2 EVIDENTIAL REASONING AND BIOINFORMATICS	9/17: - Intro belief functions (BFs) & evidential reasoning (ER) - Differences between traditional probabilistic methods and BF-ER-based methods 9/19: - Making decisions in bioinformatics. - BF-ER-based decision making.	Learning Module #2 The deadline for students to Add or Drop classes via MySJSU with no petition is September 17 (per University Policy S22-6)
Week 6	9/24	9/26	#2 EVIDENTIAL REASONING AND BIOINFORMATICS	9/24: - Computational implementation of Dempster's Rule (DR) with respect to ER. 9/26: - **Quiz 1 (~45 mins): Covers Topics From Week 1 thru Week 5**	Learning Module #2
Week 7	10/1	10/3	#2 EVIDENTIAL REASONING AND BIOINFORMATICS	10/1: - Implementing ER and DR. 10/3: - **In-Class Exercise 2 Covers Topics From In-Class Exercise 1 to Week 6**	Learning Module #2
Week 8	10/8	10/10	#2 EVIDENTIAL REASONING AND BIOINFORMATICS	10/8: - Interpreting ER results. - Example applications of BF and ER 10/10: - Example applications of BF and ER	Learning Module #2

Week 9	10/15	10/17	#3 ADVANCED HMMs	10/15: - Brief review of traditional HMMs 10/17: - **Midterm (Full period): Covers Topics From Week 1 thru Week 8**	Learning Module #2
Week 10	10/22	10/24	#3 ADVANCED HMMs	10/22: - Profile and jump HMMs 10/24: - Pair-HMMs, sub-HMMs,	Learning Module #2
Week 11	10/29	10/31	#3 ADVANCED HMMs	10/29: - Phylo-HMMs 10/31: - Examples of profile and pair HMMs	Learning Module #2
Week 12	11/5	11/7	#3 ADVANCED HMMs	11/5: - Examples of profile and pair HMMs 11/7: - **Quiz 2 (~45 mins): Covers Topics From Quiz 1 thru Week 11**	Learning Module #2
Week 13	11/12	11/14	#4 EPIGENETICS	11/12: - Intro to epigenetics 11/14: - **In-Class Exercise 3 Covers Topics From In-Class Exercise 2 to Week 12**	Learning Module #3

Week 14	11/19	11/21	#4 EPIGENETICS	11/19: - Types of epigenetic modifications - Epigenetic DBs 11/21: - Epigenetic DB APIs	Learning Module #3
Week 15	11/26	11/28	#4 EPIGENETICS	11/26: - **In-Class Exercise 4 (Work on Projects, Q&A)** 11/28: **THANKSGIVING HOLIDAY**	Learning Module #3
Week 16	12/3	12/5	#4 EPIGENETICS	12/3: - Evaluating epigenetic analysis results 12/5: - **Quiz 3 (~45 mins): Covers Topics From Quiz 2 thru Week 16**	Learning Module #3
			Final Project Report and Code Due To Canvas Wednesday December 11, 2024 By 11:59PM **No Final Exam.** **The Project Takes The Place Of The Final Exam**		

SCHEDULE FOOTNOTES:
NONE AS OF AUGUST 2024

Grades *

WRITTEN HOMEWORK (3 at 10 points each)	30 pts
PROGRAMMING ASSIGNMENTS (4 at 50 points each)	200 pts
QUIZZES (3 at 50pts each)	150 pts
MIDTERM	100 pts
IN-CLASS EXERCISES (4 at 50pts each)	200 pts
WEEKLY COURSE FEEDBACK (14 at 5pts each)	70 pts
PROJECT REPORT & CODE (200pts)	200 pts
Total Course Points	= 1,050 pts Total

* The total points for each category might change depending on the number of project teams and assignments. The instructor reserves the right to adjust, with sufficient advanced notice, the above point distributionby±5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Grading Percentage Breakdown		
Percent of Total Points	**Points**	**Letter Grade**
96.66%	≥ 1015	A plus
93.33%	≥ 980	A
90.00%	≥ 945	A minus
86.66%	≥ 910	B plus
83.33%	≥ 875	B
80.00%	≥ 840	B minus
76.66%	≥ 805	C plus
73.33%	≥ 770	C
70.00%	≥ 735	C minus
66.66%	≥ 700	D plus
63.33%	≥ 665	D
60.00%	≥ 630	D minus
59.99%	< 630	F

(NOTE: Ranges might change if point totals change)

How To Calculate/Estimate Your Grade

If students would like to calculate their numeric grade percentage, the formula is as follows:

Numeric Grade Percentage =

$$\frac{Total\ points\ from\ assignments}{Total\ course\ points} x\ 100\%$$

There is no guarantee that grades will be curved. If so, it will typically be done at the end of the semester. The instructor is already aware that graduate students need to maintain an overall GPA of B or better. Just because a student NEEDS a particular grade doesn't mean that the instructor will automatically GIVE the student that grade. Students must EARN a passing grade based on submitted and evaluated course work.

Extra Credit Options

There are no pre-planned extra credit assignments in this course. However, homework assignments and exams might, on occasion, contain extra credit options/questions. At times, the instructor might announce the availability of extra exercises or assignments. There is no guarantee that such extra credit exercises or assignments will be offered to the class. If, in the opinion of the instructor, offering such extra credit options will be significantly advantageous to the learning process, they might be offered.

Late Assignment Submission

Late assignments will receive a 25% point deduction of a graded assignment for each 24hr period the submission is late. For example, if an assignment is worth 10 points, and the grade for the assignment is 8/10, and the assignment is submitted one day late, then the point deduction equals 2.5, and the final grade for the assignment is MAX(0, 8 – 2.5) = MAX(0, 5.5) = 5.5 .

Missed Assignments, In-Class Exercises, Quizzes, and Exams

A. QUIZZES:
   a. The grade for one missed quiz will be replaced with the average of the remaining two quizzes.
   b. More than one missed quiz will result in a grade of incomplete provided the total missed points for the semester is less than 20% of the total course points.

B. MIDTERM:
   a. The grade for a missed midterm exam will be 75% of the average score for quizzes, programming assignments, and homework assignments provided the total missed points for the semester is less than 20% of the total course points. Or, provide acceptable documentation of the reason for missing the midterm as described in version 1 of this course syllabus and a makeup exam will be provided.

C. HOMEWORK ASSIGNMENTS:
   a. The grade for one missed homework assignment will be replaced with the average of the remaining three homework assignments.
   b. The grade for the second missed homework assignments will be replaced with 75% of the average of the remaining two homework assignments.
   c. More than two missed homework assignments will result in a grade of incomplete provided the total missed points for the semester is less than 20% of the total course points. An alternative is to accept zeros for the missed homework assignments, or if acceptable documentation of the

reason for missing the homework assignments is provided, makeup assignments will be provided.

D. PROGRAMMING ASSIGNMENTS:
   a. The grade for one missed programming assignment will be replaced with 50% of the remaining programming assignment.
   b. Two missed programming assignments will result in a grade of incomplete provided the total missed points for the semester is less than 20% of the total course points. An alternative is to accept zeros for all missed programming assignments, or if acceptable documentation is provided, makeup assignments can be provided

E. IN-CLASS EXERCISES:
   a. The grade for one missed In-Class Exercise will be replaced with the average of the remaining three In-Class Exercises.
   b. The grade for two missed In-Class Exercises will be replaced with 75% of the average of the remaining two In-Class Exercises.
   c. More than two missed In-Class Exercises will result in a grade of incomplete provided the total missed points for the semester is less than 20% of the total course points. An alternative is to accept zeros for all missed in-class exercises, or if acceptable documentation of the reason for missing the IN-Class Exercises is provided, a makeup assignment can be provided.

F. WEEKLY FEEDBACK:
   a. All missed weekly feedback assignments will receive zero points.

G. FINAL PROJECT REPORT & CODE:
   a. The grade for a missed final project report and code will be 75% of the average of all other course assignments, exams, and quizzes provided the total missed points for all other assignments is less than 5% of the total course points.
   b. If the total missed points for all other assignments is more than 5% but less than 20% of the total course points, a grade of incomplete will be given.

H. TOTAL MISSED POINTS MORE THAN 20% BUT LESS THAN 30% OF TOTAL COURSE POINTS AND TOTAL MISSED POINTS MORE THAN 30%.
   a. **Missed between 20% and 30% of total course points:** A course grade that equal to (100% - <missed points %>) * Average of remaining assignments, quizzes, exams, and programming assignments.
   b. **Missed more than 30% of total course points:** If the percentage of total missed points is greater than 30%, a course grade that is the result of assigning a zero grade for all missed assignments will be assigned. An alternative grade or options can be discussed with the instructor.

Receiving An Incomplete (I) Grade

Receiving a grade of Incomplete (I) is not automatic. Students must complete at least 80% of course assignments by the end of the semester to be eligible to receive a grade of incomplete. Students must also provide documentation to support the reason for the request to receive an Incomplete grade. The instructor has the final decision to give an Incomplete grade. If the instructor agrees to give a student an Incomplete grade, the instructor will enter the remaining work to be completed as part of the PeopleSoft grade submission process.

Grade Change Policy

It is a university policy (S09-7) that "A change of grade request must be submitted by the department office directly to the Office of the Registrar in a timely fashion. Normally, such requests must be received by the drop deadline of the following Spring or Fall semester … Requests for exceptions to this policy must be accompanied with a documented and compelling reason. …"

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/. Make sure to review these policies and resources.