San José State University
Computer Science Department
CS 223 Bioinformatics, Sec 01, Fall 2022

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 2022 semester.)
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Tuesdays 6:30AM – 8:30AM, Zoom Link For Office Hours For Fall 2022:: https://sjsu.zoom.us/j/81579190359?pwd=RlMwUGduVy9rSUFMeVZnU2YyMTdIZz09 Passcode 355621
Class Days/Time:	Section 01: Tuesdays and Thursdays 3:00PM – 4:15AM
Classroom:	MH 422
Prerequisites:	CS/BIOL/SE 123B. Allowed Declared Major: Computer Science, Bioinformatics, Data Science.

CoS COVID-19 and Monkeypox:

Students registered for a College of Science (CoS) class with an in-person component should view the CoS COVID-19 and Monkeypox Training slides for updated CoS, SJSU, county, state and federal information and guidelines, and more information can be found on the SJSU Health Advisories website. By working together to follow these safety practices, we can keep our college safer. Failure to follow safety practice(s) outlined in the training, the SJSU Health Advisories website, or instructions from instructors, TAs or CoS Safety Staff may result in dismissal from CoS buildings, facilities or field sites. Updates will be implemented as changes occur (and posted to the same links).

Official SJSU Catalogue Course Description

The course investigates the main algorithms for solving computational problems in bioinformatics. Methods will include Hidden Markov Models for gene prediction and protein profiling, and Genetic Algorithms for biological sequence analysis and structure prediction. Students will be given programming projects.

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. At times, vast amounts of data must be analyzed and interpreted to help answer bioinformatic related problems and questions. Computational pipelines and workflows are frequently needed to process large datasets in a timely and accurate manner. Knowing how to assess time and space performance of algorithms, and knowing how to design and implement efficient algorithms for pipelines/workflows can, at times, be a requirement to address questions of interest. Example algorithms that will be investigated include Principal Component Analysis (PCA), Singular Value Decomposition (SVD) if time permits, advanced Hidden Markov Models for gene prediction or protein profiling, biological sequence analysis, and structure prediction. Students will be given programming projects that provide practice with analyzing, designing, implementing, and using bioinformatics related algorithms.

Learning Outcomes

Upon successful completion of this course, students will:

1. SLO-1 COMPUTATIONAL ANALYTICSFOR BIOINFORMATIC ALGORITHMS: Know and be able to apply key fundamental techniques in algorithm design and analysis to bioinformatic related algorithms. Topics include computational complexity (e.g., time and space) analysis methods and techniques, proof of correctness methods, matrix operations, approximation algorithms, and if time permits computational geometry. The knowledge and skills learned here will be useful

2. SLO-2 PRINCIPAL COMPONENT ANALYSIS (PCA): Know why and how PCA can be used within the bioinformatics and machine learning domains. If time permits how SVD (Singular Value Decomposition) works and the difference from PCA will be covered. Students will also apply the knowledge and skills gained from SLO-1 to PCA- and/or SVD within the context of bioinformatics-related problems.

3. SLO-3 ADVANCED HIDDEN MARKOV MODELS (HMMs): Be familiar with the time, space, and accuracy limitations of traditional HMMs. A new and evolving

evidence-based mathematics that can be used to address some of the traditional HMM limitations will be presented. Algorithm analysis techniques from SLO-1 will be applied to conduct comparative time and space complexity analyses of the traditional versus evidential HMMs.

4. SLO-4 PATTERN ANALYSIS IN BIOINFORMATICS: Some of the algorithms and methods to extract patterns from data and information to support bioinformatic analyses. Students will learn the definition of a pattern within the context of microarray gene expression analysis, protein structures, and motifs in terms of amino acids. The concept of a pattern profile will be introduced to extend this concept, allowing a quantitative description of gene expression, motifs, and by assigning probabilities to the occurrence of gene expression and a particular amino acid at each position of a motif. Algorithm analysis techniques from SLO-1 will be applied to conduct comparative time and space complexity analyses of protein pattern and profile analysis algorithms.

Each SLO above corresponds to a learning module that is described in the course calendar below. That is, there are three (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)

Bioinfromatics: 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. 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	8/23	8/25	#1 Computational Analytics (MSBI Majors) #1 Biology Basics (MSCS Majors)	**8/23:** Intro To Course: -Topics, learning objectives, course logistics, Instructor background - Syllabus 8/25: - Analyzing algorithms (MSBI) - Biology Central Dogma (MSCS)	Learning Module #1

Week 2	8/30	9/1	#1 Computational Analytics (MSBI Majors) #1 Biology Basics (MSCS Majors)	**8/30:** - Biology Central Dogma cont. (MSCS) **9/1:** - Analyzing algorithms: Sorting & Search Algs, PseudoCode, and Time+Space Complexity Analysis (MSBI)	Learning Module #1
Week 3	9/6	9/8	#1 Computational Analytics (MSBI Majors) #1 Biology Basics (MSCS Majors)	**9/6:** - Biology Central Dogma cont. (MSCS) **9/8:** - Analyzing algorithms: Asymptotic Notation, Solving Recurrence Relations. (MSBI)	Learning Module #1
Week 4	9/13	9/15	#1 Computational Analytics (MSBI Majors) #1 Biology Basics (MSCS Majors)	**9/13:** - Biology Central Dogma Bioinformatic DBs and APIs (MSCS) **9/15:** - Analyzing algorithms: Asymptotic Notation, Solving Recurrence Relations cont. (MSBI)	Learning Module #1

Week 5	9/20	9/22	#2 PCA, SVD	**9/20:** - PCA - Strassen matrix multiplication **9/22:** Module #1 **In-Class Exercise 1** - **Covers topics in Week1 thru Week 4**	Learning Module #2
Week 6	9/27	9/29	#2 PCA, SVD	**9/27:** - PCA, time & space complexity comparison and analysis. **9/29:** **Quiz 1 (~35 mins):** **Covers Topics Week 1 thru Week 5**	Learning Module #2
Week 7	10/4	10/6	#3 Advanced HMMs	**10/4:** - Advanced HMMs - Analysis of HMMs **10/6:** **In-Class Exercise 2** **Covers topics in Week 4 thru 6**	Learning Module #3
Week 8	10/11	10/13	#3 Advanced HMMs	**10/11:** - Analysis of HMMs - Evidential Reasoning **10/13:** - Evidential Reasoning and HMMs	Learning Module #3

Week 9	10/18	10/20	#3 Advanced HMMs	**10/18:** - Evidential Reasoning and HMMs **10/20:** **Midterm (Full period): Covers Topics Week 1 thru Week 7**	Learning Module #3
Week 10	10/25	10/27	#3 Advanced HMMs	**10/25:** - Evidential Reasoning and HMMs **10/27:** - Finish Evidential Reasoning and HMMs	Learning Module #3
Week 11	11/1	11/3	#4 Pattern Analysis	**11/1:** - Intro Pattern Analysis **11/3:** - Pattern Analysis Profiles	Learning Module #4
Week 12	11/8	11/10	#4 Pattern Analysis	**11/8:** **In-Class Exercise 3 Topics Coverd Week 8 to Week 11** **11/10:** **Quiz 2 (~40 mins): Covers Topics Week 8 thru Week 12**	Learning Module #4
Week 13	11/15	11/17	#4 Pattern Analysis	**11/15:** - Pattern Analysis Moifs **11/17:**	Learning Module #4

Week 14	11/22	11/24	#4 Pattern Analysis	**11/22:** - Pattern Analysis Gene Expression **11/24:** **THANKSGIVING**	Learning Module #4
Week 15	11/29	12/1	#4 Pattern Analysis	**11/29:** - Pattern Analysis motifs **12/1:** **In-Class Exercise 4** (work on Topic Project)	Learning Module #4
Week 16	12/6	N/A	#4 Pattern Analysis	**12/6:** **Quiz 3 (~35 mins):** **Covers Topics Week 9 thru Week 13**	Learning Module #4
			Final Project Report and Code Due To Canvas **Wednesday December 14, 2022 By 11:59PM**		

SCHEDULE FOOTNOTES:
NONE AS OF AUGUST 2022

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 .

Making Up Missed Assignments

An opportunity to makeup missed exams, homework, in-class exercises, programming assignments, and so forth will be provided if and only if verifiable documentation of a compelling reason (e.g., illness, COVID-19 infection, accident, death in the immediate family, work related) for missing the assignment is provided within one week from the student's ability to return to class. It is the student's responsibility to (1) contact the instructor if an assignment has or will be missed; (2) obtain verification from the instructor that the student will be allowed to make up the assignment, subject to acceptable and verified documentation; and (3) make arrangements with the instructor to submit all missing assignments by the sooner of a date specified by the instructor or by the end of the semester, whichever is earlier. Failure to provide verifiable documentation of the reason for missing the assignment or failure to provided such within one week from the student's ability to return to class will result in not being able to make up the missed assignment.

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.