San José State University
Computer Science Department
CS 223: Bioinformatics,
Section 1, Spring 2017

Course and Contact Information

Instructor:	Leonard Wesley
Office Location:	MacQuarrie Hall 212 (MH 212)
Telephone:	408.924.5287
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Tuesdays 2:00PM – 4:00PM
Class Days/Time:	Tuesdays and Thursdays Sec 1: 7:30PM – 8:45PM
Classroom:	MH 422
Prerequisites:	Prerequisite: CS 123A or CS 155
GE/SJSU Studies Category:	N/A

Course Format

Faculty Web Page and MYSJSU Messaging (Optional)

Course materials such as syllabus, handouts, notes, slides, assignment instructions, and so forth can be found on the Canvas Course Learning Management System for this course. You are responsible for regularly checking, at least 3 times per week, with the messaging system through MySJSU at http://my.sjsu.edu (or other communication system as indicated by the instructor) to learn of any updates. Email communication from the instructor to the class will be sent primarily from Canvas. To make sure students receive such communication, students should specify, within the Canvas shell, forwarding of Canvas initiated email to whatever email address they prefer and is checked regularly.

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.

Course Learning Outcomes (CLO) (Required)

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

1. SLO 1 biology basics such as the central dogma, DNA transcription, translation, protein synthesis and tge protein four structure levels. Students will also be familiar with genomic sequencing and genome assembly methods.
2. SLO 2 the theory and application of HMMs, genetic algorithms, sequence alignment, clustering methods, greedy algorithms, probabilistic models and algorithms, Monte Carlo simulation, and ANNs to carry out tasks such as gene identification, biomarker identification, motif finding, protein structure prediction & scoring, and genomic alignment.
3. SLO3 how to use BioPython and scikit's Python-based ML modules to help achieve SLO 2

Required Texts/Readings

Textbook

MACHINE LEARNING APPROACHES TO BIOINFORMATICS, Zheng Rong Yang , World Scientific Publishing Co. Pte. Ltd. 2010, ISBN-13 978-981-4287-30-2, ISBN-10 981-4287-30-X

Other Readings

Bioinformatics: The Machine Learning Approach, Pierre Baldi and Soren Brunak, 2nd Edition, MIT Press, 2001 ISBN 0-262-02506-X

ADVANCED ANALYSIS OF GENE EXPRESSION MICROARRAY DATA, Aidong Zhang , World Scientific Publishing Co. Pte. Ltd., 2006 , ISBN 981-256-645-7

The instructor will likely provide additional reading, study, and exercise materials, as well as provide web-based links to relevant material as appropriate.

Other technology requirements / equipment / material

Students are required to have a working Python 2.7.x or 3.x environment (preferred is Enthought Canopy Python (see https://www.enthought.com/products/canopy/). The instructor will be using the latest version of Enthought Canopy Python 2.7.x with all updates installed. Programming assignments will be graded based on execution in the instructor's environment and not on whether students are able to run their submissions on other platforms. Therefore it is highly advisable that students install and used the Enthought Canopy Python

2.7.x environment. Instruction and assignments will be given using Python and submissions of assignments and project code will be expected to be in Enthought Canopy Python 2.7.x.

Course Requirements and Assignments

SJSU classes are designed such that in order to be successful, it is expected that students will spend a minimum of forty-five hours for each unit of credit (normally three hours per unit per week), including preparing for class, participating in course activities, completing assignments, and so on. More details about student workload can be found in University Policy S12-3 at http://www.sjsu.edu/senate/docs/S12-3.pdf.

Course Logistics

Students will be assigned reading, possibly video, homework, programming, and related assignments on a near weekly basis. Periodically, as per the Greensheet, students will be required to review and complete in-class exercises associated with each topic area covered in the course. Teams of approximately three to four students will be formed to work on and complete in-class exercises. Quizzes, exams, homework, and programming exercises will be assigned as specified below in the section entitled "**Tentative course calendar of assignment…"** below.

Students should expect to spend approximately eight (8) to 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, programming, completing homework and programming exercises, and so forth. The amount of time that a student actually spends will depend on individual skills and the time allocated to the course. The eight (8) to nine (9) hours per week estimate is based on previous experiences of the instructor and students. So please plan and schedule accordingly. It is also suggested that students complete work that is outside of the classroom in a minimum of 3hr blocks of time. It is typically more difficult and less efficient to complete programming assignments in much smaller (e.g., 30min or 1hr) blocks of time.

Lecture presentation material will consist, in part, of a combination of not only lecture slides, but verbal and white board presentation lectures where it is recommended that students always be prepared and ready to take notes of presented material and instructions regarding assignments and so forth.

Previously, students have asked for special exception 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 view of the instructor, no exception will be given to a student unless it 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 available to the entire class.

NOTE that University policy F69-24 at http://www.sjsu.edu/senate/docs/F69-24.pdf 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.

Final Examination or Evaluation, Quizzes, Exams, Assignments, and In-Class Exercises

There will be three quizzes, one midterm, four in-class exercises, four programming assignments, and one final exam (or project report and code submission) that will count toward the final grade (percentage wise) 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. Copying or duplicating the work of others is also strictly prohibited without the express permission of the instructor. Violations will result in a grade of zero for the assignment, exam, assignment, and will be met with the full impact of SJSU's academic integrity policy and procedures.

Missed Exams and Quizzes

There are no makeup exams or makeup quizzes. If the Midterm is missed, the average of all quizzes will be substituted for the missed Midterm exam. If one quiz is missed, the average of the remaining two quizzes will be substituted for the missed quiz. If two quizzes are missed, the average of the remaining quiz score and a zero (0) score will be substituted for both missed quizzes. If all three quizzes are missed, then a score of zero (0) will be assigned to each quiz and counted toward the final grade.

Missed and Late Programming Assignments and In-Class Exercises

Missed Programming and In-class exercises will receive a grade of zero.
Late programming assignments will have 20% of the max points deducted for each day the assignment is late. For example, if the max points for a programming assignment is 100pts, and the assignment is two days late, and the score for the submission is 90/100 then 40pts will be deducted from the score of 90 resulting in and a grade of 50 being recorded for the assignment. There are no late in-class assignments.

Projects

Several AI-related project topics will be described near the start of the course. Projects will involve applying one or more of the skills and knowledge learned in the course to the project. Teams of 3-4 students will be formed to work on a selected project topic. Teams will be required to submit a project proposal before starting on a project, and give a project presentation at the end of the course. Individual student scores on a project will be determined by the content and quality of the contribution of each student toward the project. The score on the course project and project presentation will count toward the final grad (percentage wise) as specified in the "Grades" section below.

Completing and submitting a final course project report and code is mandatory to receive a grade in the course. Failing to submit a project and code will result in a grade of F.

Reading and Homework Assignments
Students might be assigned videos to view and/or complete assigned textbook readings or programming assignments as specified in the course calendar of assignments shown below. Homework assignments will be assigned as indicated in the Greensheet.

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.

Grading Information*

QUIZZES (3 each 50pts)	150 pts
MIDTERM	200 pts
IN-CLASS EXERCISES (4 @ 25pts each)	100 pts
PROGRAMMING ASSIGNMENTS (4 @ 50pts each)	200 pts
HOMEWORK ASSIGNMENTS (4 @ 25pts each)	100 pts
FINAL EXAM/PROJECT RPT & CODE	200 pts
PRESENTATION & PRESENTATION PARTICIPATION	50 pts
Total Course Points	= 1,000 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 distribution by ± 5 pts. Such adjustments might be based on the difficulty or simplicity of assignments or quizzes or exams.

Determination of Grades

Grading Percentage Breakdown (NOTE: Ranges might change if point totals change)

Percentage of Total Pts	Pts	Letter Grade
98.00% and above	> 980	A+
93.33% - 97.99%	933 - 979	A
90% - 93.32%	900 – 932	A-
86.66% - 89.99%	866 – 899	B+
83.33% - 86.65%	833 – 865	B
80% - 83.32%	800 – 832	B-
76.66% - 79.99%	766 – 799	C+
73.33% - 76.65%	733 - 765	C
70% - 73.32%	700 – 732	C-
66.66% - 69.99%	666 – 699	D+
63.33% - 66.65%	633 – 665	D
60% - 63.32%	600 - 632	D-
Below 60%	**< 600**	**F**

HOW TO CALCULATE/ESTIMATE YOUR GRADE

If students would like to calculate their numeric grade percentage, the formula is as follows:
Numeric CS 223 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 be done at the end of the semester. The instructor is already aware that graduate students need to maintain an overall GPA of 3.0 or better, and that undergraduates need to maintain an overall GPA of 2.0 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. Students should be aware that achieving a particular grade in another course will not influence your grade for this course.

All questions on Homework Assignments and all aspects/portions of code of a Programming Assignment might not be evaluated. On homework assignments, at times just one or two questions will be selected, at the instructors discretion and unknown to students, for grading. The entire assignment will receive the grade of the score for whatever questions were selected for grading. For example, suppose a homework assignment has four questions for a total of 100pts. If, for instance, question 1 is graded and worth 20 pts, and 5pts were deducted from the first question. 15/20 pts x 100% = 75%. Then a score of 75% or 75 points will be assigned for the entire homework assignment. The question(s) actually graded will not be announced to the class until after the assignment is graded.

Extra credit options, if available
There are no extra credit assignments in this course except for completing designated "Advanced" assignments. However, homework assignments and exams might contain extra credit options.

Penalty for late or missed work
Late assignments will not be graded and receive a score of zero. There will be partial credit for assignments. If you cannot complete an assignment by a given deadline, then submit what you have completed to be eligible to receive partial credit. Missed assignments (written or programming) will receive a grade of zero.

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 that course grade changes must be made within one semester from the end of the course. Requests for exception must be accompanied with compelling and verifiable documentation.

Classroom Protocol

Exams and quizzes will start approximately five (5) minutes after the official class start time. Once the quiz and exam starts, the door will be closed and students will not be allowed in until the quiz or exam is over. This is to preclude late arrivals from disrupting students that have already started the quiz or exam. Once a quiz or exam begins and a student leaves before the quiz or exam is completed, re-entry to the classroom and continuing with the exam will not be allowed. Therefore, it is advised that students go to restrooms before the quiz or exam begins.

Cell Phones and Pagers:
Students will turn their cell phones and pagers off or put them on vibrate mode while in class. Students will not be allowed to answer or text with their phones or pagers in class. Students whose phones or pagers disrupt the class and do not stop when requested by the instructor will be referred to the Judicial Affairs Officer of the University.

Computer Use:
In the classroom, students who use their computers in a disruptive way (e.g., clicking of key board is too loud) or who abuse the classroom equipment in any way, at a minimum, will be asked to leave the class and will lose participation or in-class exercise points for the day, and, at a maximum, will be referred to the Judicial Affairs Officer of the University for disrupting the course. (Such referral can lead to suspension from the University.) Students are urged to report to their instructors computer use that they regard as inappropriate (i.e., used for activities that are not class related).

Right to Privacy:
The student will retain a right to privacy. The instructor will not knowingly reveal students' grades, student ID numbers, phone numbers, addresses or other private information to others, except within the limits of university policy. You will be asked to supply your first name, last name and last four digits of your SID on quizzes and exams.

Consent for Recording of Class and Public Sharing of Instructor Material
University Policy S 12-7, htlp://www.sjsu.edu/senate/docs/S12-7.pdf requires students to obtain the instructor's permission to record the course. Common courtesy and professional behavior dictate that you notify someone when you are recording them. You must obtain the instructor's, permission to make audio or video recordings in this class. Such permission allows the recordings to be used only for your private study purpose. The recordings are the sole intellectual property of the instructor. You have not been given any right, nor do you have any explicit or implicit right to reproduce or distribute the course material developed by the instructor. Course material developed by the instructor is the intellectual property of the instructor and cannot be shared publicly without the instructor's approval. You may not share or upload instructor generated material for this course such as exam or quiz questions, lecture notes, in-class or hands-on exercises, or solutions without the instructor's explicit written permission.

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/"

CS223 Bioinformatics, Spring 2017, Course Schedule

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")

Course Schedule

Week	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
1	1/24 No Class	1/26 First Class	1/26: Introduction To Course Review Greensheet Prerequisite Check Exam Start Introduction To Bioinformatics ➢ What is bioinformatics ➢ Cell structure, DNA, RNA, …, Proteins , Central Dogma	Week 1 Assignment 1/26: Read Biology Review Handout
2	1/31	2/2	1/31: Continue Introduction To Bioinformatics ➢ Eukaryotes & Prokaryotes ➢ Genome sequencing, sequence assembly, sequence alignment 2/2: Continue Introduction To Bioinformatics ➢ Homologs, Orthologs, Paralogs ➢ SNPs, Motifs,	Week 2 Assignment 1/31 and 2/2: Read Biology Review Handout
3	2/7	2/9	2/7: Finish Introduction To Bioinformatics ➢ Example use of various ML algorithms to help identify genes, biomarkers, motifs, alignments, …etc. 2/9: Intro To Hidden Markov Models (HMMs) and gene finding.	Week 3 Assignment **Programming Assignment 1** **Homework Assignment 1** 2/7: HMM Reading Assignment 2/9: HMM Reading Assignment
4	2/14	2/16	2/14: Discussion Of Possible Projects HMMs continued 2/16: **In-Class Exercise 1** HMMs	Week 4 Assignment 2/14: Prep For In-Class Exercise

Week	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
				2/16: Finish In-Class Exercise
5	2/21	2/23	2/21: **Quiz 1:** Covers topics from 1/26 to 2/16 2/23: Intro To Genetic Algorithms and protein sequence alignment, and clustering	Week 5 Assignment 2/23: Genetic Algorithm Reading Assignment
6	2/28	3/2	2/28: Genetic Algorithms cont. 3/3: Genetic Algorithms cont.	Week 6 Assignment **Programming Assignment 2** **Homework Assignment 2** 2/28: Genetic Algorithm Reading Assignment 3/3: Prep For In-Class Exercise
7	3/7	3/9	3/7: **In-Class Exercise 2** Genetic Algorithms 3/9: Intro To Probabilistic & Evidential Models, Methods, Inference, and application to biomarker identification, gene finding, or protein structure formation.	Week 7 Assignment 3/9: Probabilistic & Evidential Reading Assignment
8	3/14	3/16	3/14: Intro To Probabilistic & Evidential Models, Methods, Inference … cont. 3/16: Intro To Probabilistic & Evidential Models, Methods, Inference … cont.	Week 8 Assignment 3/14: Probabilistic & Evidential Reading Assignment 3/16: Probabilistic & Evidential Reading Assignment

Week	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
9	3/21	3/23	3/21: Finish Intro To Probabilistic & Evidential Models, Methods, Inference … cont. 3/23: **Midterm** Covers topics from 1/26 to 3/21	Week 9 Assignment **Programming Assignment 3** **Homework Assignment 3**
			SPRING RECESS 3/27 – 3/31	
10	4/4	4/6	4/4: Artificial Neural Networks (ANNs) and application to gene finding, microarray image analysis 4/6: ANNs cont.	Week 10 Assignment 4/4: ANN Reading Assignment 4/6: ANN Reading Assignment
11	4/11	4/13	4/11: ANNs cont. 4/13: **Quiz 2** Covers topics from Midterm to 4/11.	Week 11 Assignment 4/13: Prep For In-Class Exercise
12	4/18	4/20	4/18: **In-Class Exercise 3** ANNs 4/20: Monte Carlo Simulation and application to spatial biology such as modeling of ecological multi-species cross-feeding.	Week 12 Assignment **Programming Assignment 4** **Homework Assignment 4** 4/20: Monte Carlo Simulation Reading Assignment

Week	Tue	Thur	SUBJECT/TOPIC	Assignment (See Canvas For Due Dates)
13	4/25	4/27	4/25: Monte Carlo Simulation … cont. 4/27: **Quiz 3** Covers topics from Midterm to 4/25	Week 13 Assignment 4/2-275: Monte Carlo Simulation Reading Assignment
14	5/2	5/4	5/2 : Monte Carlo Simulation … cont. 5/4: **In-Class Exercise 4** Monte Carlo Simulation	5/2: Prep For In-Class Exercise
15	5/9	5/11	5/9: Work on Project 5/11: Project Presentations	None
16	5/16	5/18 No Class	5/16: Project Presentations	None
Final Project Report & Code Due Date and Time Is Specified In The Course Canvas Shell The course final project & code IS the course final exam				