San José State University
Computer Science Department
CS 286: Advanced Topics In Computer Science, Sec 02
Bioinformatics Computational Analytics, Spring 2016

1. Course Information

Instructor: Leonard P. Wesley

Department: Computer Science
College of Science, San Jose State University.

Fall Semester, 2015

Course and Contact Information

Instructor:	Leonard Wesley
Office Location:	MH 212
Telephone:	408.924.5287
Email:	Leonard.Wesley@sjsu.edu
Office Hours:	Tuesdays and Thursdays 1:00PM – 2:00PM
Class Days/Time:	Tuesdays and Thursdays 6:00PM – 7:15PM
Classroom:	DH 450
Prerequisites:	Graduate standing or instructor consent.

Course Description

Students will learn how to analyze and characterize the accuracy, precision, completeness, speed, and memory usage of computational algorithms used to address problems related to large quantities of biological data. The algorithms characterized are typically used in Genome Assembly, Meta-genomics, Microarrray analysis, Drug Discovery and Development. Additional application domains include business intelligence, security, and data analytics. Example algorithms include (1) Dimension Reduction (e.g., PCA, Factor Analysis, …) (2) generalized de Bruijn Digraphs, (3) Linear SVM, (4) HMMs, (5) Clustering (e.g., k-means, hierarchical), and others as time permits.

Learning Outcomes

Upon successful completion of this course, students will become familiar with:

1. several performance aspects such as accuracy, speed, completeness, precision, and memory usage of selected algorithms and methods, such as HMMs, gene expression analysis, genome assembly, SVM, and related methods that are used in the bioinformatics, life science, and related domains of application.
2. the methods and techniques that can be used to help identify, characterize, and interpret several metrics that can be used to help assess the above performance aspects.
3. a few domains and case studies to which the above course learning outcomes can be applied.

Required Texts

Reading handouts will be provided. Additional required reading material will be distributed to the class as appropriate.

Other Optional Reading Material

Optional reading material will be provided as appropriate.

Course Logistics

Teams of approximately three students will be formed. Each team will be assigned a topic paper to read in depth and lead a class discussion and presentation about the topic. For each topic, each team will be given a set of questions, exercises, and examples that are to be presented and discussed during the class presentation. Every student that is not presenting will be required to read the assigned paper(s) before presentations and submit a completed assignment related to that topic and current presentation.

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 so forth. The amount of time that a student actually spends depends on their individual skills and the time allocated to the course. The nine (9) hours per week estimate is based on the previous experiences of the instructor and students. So please plan and schedule accordingly.

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 them to pursue.

Quizzes and Exams

There will be one midterm and one final exam that will count toward the final grad (percentage wise) as specified in the "Grades" section below.

Projects

Several bioinformatics-related, life science, and/or big data related course project topics will be described near the start of the course. Projrects 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.

Computational Resources

Students are required to make sure that they have access to sufficient UNIX or Windows 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.

4. Tentative course calendar of assignment due dates & exam dates:

(Please note that course calendar & content is "subject to change with fair notice")

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment
1	No Class	1/28	Introduction To Computational Analytics Computational Analytics in both the biological and big data domains. Application Areas Organization of each class lecture (using biological and/or big data context) Course Learning Objectives What is Dimension Reduction & Why Is It Needed To Improve Accuracy & Performance? Examples from Cancer Related Gene Identification, Drug Candidate Prediction. Start Introduction of Dimension Reduction Methods.	Dimension Reduction: Reading Assignment #1 Homework Assignment #1
2	2/2	2/4	History/Background of Data Analytics -> Computational Analytics Continuation of Dimension Reduction Methods. Impact on Performance & Accuracy	
3	2/9	2/11	Continuation of Dimension Reduction Methods. Impact on Performance & Accuracy, Selecting the appropriate dimension reduction methods, tradeoffs.	
4	2/16	2/18	Continuation of Dimension Reduction Methods. Impact on Performance & Accuracy, Selecting the appropriate dimension reduction methods, tradeoffs.	
5	2/23	2/25	Continuation of Dimension Reduction Methods. Impact on Performance & Accuracy, Selecting the appropriate dimension reduction methods, tradeoffs. Project Ideas Discussion	
6	3/1	3/3	Methods For Sequence Assembly De BruijnDigraphs • Examples continued from bioinformatics (assembling bacterial and other microbial genomes) • Computational Considerations.	

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment
7	3/8	3/10	**MIDTERM**	
8	3/15	3/17	Methods For Sequence Assembly Continued De BruijnDigraphs Continued • Examples continued from bioinformatics (assembling bacterial and other microbial genomes) • Computational Considerations. • Selecting the appropriate alg/method/model & tradeoffs	
9	3/22	3/25	Introduction To Linear Support Vector Machines (SVMs) • Examples from bioinformatics (protein ligand affinity & activity prediction) • Computational Considerations. • P >> N and dimension reduction.	
			SPRING RECESS 3/28 – 4/1	
10	4/5	4/7	Linear Support Vector Machines (SVMs) Continued PCA Analysis • Examples from bioinformatics (protein ligand affinity & activity prediction) • Computational Considerations. • P >> N and dimension reduction. • Selecting the appropriate alg/method/model & tradeoffs. Examples from bioinformatics (phylogeny trees)	
11	4/12	4/14	Linear Support Vector Machines (SVMs) Continued PCA Analysis ➢ Examples from bioinformatics (protein ligand affinity & activity prediction) & big data (facial recognition). ➢ Computational Considerations.	

Mtg #	Tue	Thur	SUBJECT/TOPIC	Assignment
			o P >> N and dimension reduction. ➢ Selecting the appropriate alg/method/model & tradeoffs. Examples from bioinformatics (phylogeny trees) ➢ Computational Considerations. ➢ Selecting the appropriate alg/method/model & tradeoffs	
12	4/19	4/21	Introduction to HMMs ➢ Examples from bioinformatics (Genomic or proteomic sequence matching & alignment) Computational Considerations. o Scale issues	
13	4/26	4/28	HMMs Continued ➢ Examples from bioinformatics (Genomic or proteomic sequence matching & alignment) ➢ Computational Considerations. o Scale issues	
14	5/3	5/5	Introduction To Clustering Methods, Performance, and Accuracy k-means, hierarchical	
15	5/10	5/12	Project Presentations	
Final Project Report Due Date and Time Is Specified In The Course Canvas Shell **Final Exam Thursday May 19 2016 5:15PM – 7:30PM DH 450**				

SCHEDULE FOOTNOTES: (NONE as of Spring 2016)

NONE AS OF JAN 28, 2016

Grades *

Group Topic Presentations: likely 2 (100 pts each) x 2 = 200 pts

Midterm = 300 pts

Homework: likely 6 (lowest score dropped) (50 pts each) x 6 = 60 pts

Final Proj. Report & Presentation (Rpt 200pts, Presentation 100) = 300 pts

Final Exam = 400 pts

Total Course Points = 1,260 pts Total

* The total points for each category might change depending on the number of project teams and homework 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 exams.

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

Percentage of Total Pts	**Pts**	**Letter Grade**
96.66% and above	> 1217	A+
93.33% - 96.65%	1136-1216	A
90% - 93.32%	1022-1135	A-
86.66% - 89.99%	886-1021	B+
83.33% - 86.65%	738-885	B
80% - 83.32%	590-737	B-
76.66% - 79.99%	452-589	C+
73.33% - 76.65%	331-451	C
70% - 73.32%	232-330	C-
66.66% - 69.99%	155-231	D+
63.33% - 66.65%	98-154	D
60% - 63.32%	59-97	D-
Below 60%	**< 59**	**F**

HOW TO CALCULATE/ESTIMATE YOUR GRADE

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

Numeric CS 286 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 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, 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

Missed assignments (written or programming) will receive a grade of zero. There are no make-up exams or assignments.

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 exceptions to this policy must be accompanied with a documented and compelling reason.

University, College, or Department Policy Information:

a) Academic integrity statement (from Office of Judicial Affairs):

Your own commitment to learning, as evidenced by your enrollment at San José State University and the University's Academic Integrity Policy requires you to be honest in all your academic course work. Faculty are required to report all infractions to the Office of Judicial Affairs. The policy on academic integrity can be found at http://www2.sjsu.edu/senate/S04-12.pdf

Any student or students involved in a cheating incident on any quiz, midterm exam, a programming exam or the final exam will receive an F in the course, and will be reported to the judicial affairs office with a recommendation for disciplinary action.

The instructor will personally notify you of any such findings or actions. All such reports will also be brought to the attention of the Computer Science Department office. Any student accused of cheating has certain rights of appeal, which may serve to exonerate you.

b) Campus policy in compliance with the Americans with Disabilities Act:

"If you need course adaptations or accommodations because of a disability, or if you need special arrangements in case the building must be evacuated, please make an appointment with me as soon as possible, or see me during office hours. Presidential Directive 97-03 requires that students with disabilities register with DRC to establish a record of their disability."

c) Additional Policies and Procedures:

To ensure that every student, current and future, who takes courses in the College Of Science, has the opportunity to experience an environment that is safe, attractive, and otherwise conducive to learning, the instructor of COMPUTER SCIENCE 223 has established the following policies:

Cell Phones:

Students will turn their cell phones off or put them on vibrate mode while in class. They will not answer their phones in class. Students whose phones disrupt the course 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, faculty allow students to use computers only for class-related activities. These include activities such as taking notes on the lecture underway, following the lecture on Web-based PowerPoint slides that the instructor has posted, and finding Web sites to which the instructor directs students at the time of the lecture. Students who use their computers for other activities or who abuse the equipment in any way, at a minimum, will be asked to leave the class and will lose participation 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).

d) 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.

Vision

To be a learning community that empowers its students to better the world through innovative applications of science knowledge and skills.

Better the World: To achieve this vision, we intend to implement programs that will provide students an understanding of the social and economic context in which technologies are developed and used. Further, students also need to gain a firm ethical grounding, and guidance for their beneficial applications. The applications could be for social benefit, economic advancement, security, or the environmental sustainability of the world. In particular, our students need to understand the economic forces that shape the role of American engineers in today's competitive global economy.

Science Knowledge and Skills: Science majors develop their capabilities based on scientific knowledge and analytical methods. Our students need to acquire a solid foundation in the knowledge and methods that will prepare them for life-long learning in today's rapidly advancing world of technology. Further, in order to be competitive, our students must have superior knowledge of engineering and science theory and honed skills in the application of theory-to-practice. They need to master science and engineering topics that correspond to industry issues and trends as well as evolving global requirements.

Innovative Applications: In addition to learning science theory and skills, our students must have opportunities to learn innovation—a capability highly valued in today's global economy. Given its close ties to Silicon Valley industry, the College is in a unique position to focus its efforts on developing innovative applications of technologies. Innovation, defined as the development and exercise of creative processes to "see" beyond limits and boundaries, has the entrepreneurial quality of understanding and meeting customers' needs. It often occurs across disciplinary boundaries with contributing members having various functional expertise. Further, the ability to innovate contributes directly to the success of enterprises.

In summary, our vision for the College articulates our aspiration to inspire and educate our students to develop science capabilities as well as to understand the context in which such capabilities are used with the end goal of benefiting humanity.

Mission

The College of Science will provide empowering educational opportunities to students for their technical, professional and social development in a competitive and dynamic global society. We will build a vibrant community of students, faculty, staff, alumni, and industry professionals through strategic collaborations with Silicon Valley, California, national and global partners.

APPENDIX:

• In addition to my specifically posted office hours, I am available by arrangement.

• You are responsible for understanding the policies and procedures about add/drops, academic renewal, withdrawal, etc. found at http://www2.sjsu.edu/senate/S04-12.pdf

• Expectations about classroom behavior; see Academic Senate Policy S90-5 on Student Rights and Responsibilities.

• As appropriate to your particular class, a definition of plagiarism, such as that found on Judicial Affairs website at http://www2.sjsu.edu/senate/plagarismpolicies.htm

• If you would like to include in your paper any material you have submitted, or plan to submit, for another class, please note that SJSU's Academic Integrity policy S04-12 requires approval by instructors."