MCB 447 – Big Data in Molecular Biology and Biomedicine

SYLLABUS – Fall 2019

Location and time: TBA

Description of Course

Recent technological advances enable the collection of massive biological data sets, both in the research lab and in the medical clinic. These Big Data offer opportunities for discovering new biology, but they also demand new analysis approaches. This course will introduce students with a strong molecular biology background to the use of Big Data statistics. Students will learn how to visualize complex data, identify biologically relevant clusters, model relationships between variables, and classify entities. They will apply these techniques to a diverse range of biological Big Data, including electronic medical records, gene expression measurements, and human population genetic sequences. Students will learn through homework, in-class exercises, and a substantive final project.

Course Prerequisites

MCB 181 (Introductory Biology I)
Math 263 (Introduction to Statistics and Biostatistics) or equivalent coursework
Math 122A/B or 125 (Calculus I) or Math 123A (Mathematics of Biological Systems)

Instructor and Contact Information

Prof. Ryan Gutenkunst, PhD
Life Sciences South 325
rgutenk@email.arizona.edu

Office hours: Thursdays at 2:00 pm

The website for the course is on d2l, https://d2l.arizona.edu/.

Course Objectives and Expected Learning Outcomes

In this course, students will:

- Learn data analysis techniques applicable to a wide range of biological Big Data, including visualization, clustering, regression, and classification
- Apply those techniques to a wide variety of biological Big Data in class and as homework
- Reproduce results from a scientific paper that applies Big Data techniques to molecular biology or biomedicine

After completing this class, students will be able to:

- Evaluate the suitability of various analysis approaches for different biological questions and data
- Use the programming language R to analyze biological data
- Visualize high-dimensional data and identify biologically-relevant clusters
- Model the relationship between a given outcome and multiple potential explanatory factors, including identifying the key factors
- Classify new data points using statistical models trained from existing data
- Evaluate the use of Big Data techniques in the biological scientific literature

Required Texts or Readings

Statistical Modeling and Machine Learning for Molecular Biology by Alan Moses
I suggest you purchase the text, although the library has a handful of electronic copies.

Assignments

Homework will be assigned each Friday and due the following Friday. You are encouraged to work together and to discuss the homework assignments with your classmates. All submitted work,

however, must be your own. Publications or online sources must be also cited.

Much of class time will be spent on **in-class exercises**, often working in groups. If you are unable to finish the exercise in class, it will be due the following week. You are expected to ensure that all members of your group are following the work that is done. If you complete an assignment early, you are asked to join another group that has not finished and help out. Explaining your approach to others is an excellent way to learn material deeply.

Extra credit assignments will not be offered.

Final Project

There will be no final examination. The final project for the course will center on a scientific article that leverages big data in molecular biology or medicine, chosen from a collection of articles compiled by the instructor. Working in groups, you will give a **15-minute presentation** to your classmates explaining the article and relevant background. Also as a group, you will **reproduce the key results** of the paper. Working individually, you will lastly turn in a **5-8 page paper** summarizing the article, including background, methods, and results.

The final project presentations will be during the last week of the course, and the final paper will be due during finals week.

Grading Scale and Policies

Final scores for the course will be calculated as below, and letter grades will be assigned based on the rubric below. If necessary, the instructor may lower the curve to shift the grade distribution upward.

Homework	40%
In-class exercises	25%
Final project paper	20%
Final project presentation	10%
Final project code	5%

A	>90%
B	80-90%
C	70-80%
D	60-70%
E	<60%

Late work policy: Work less than one week late will have a 25% score reduction. Work more than one week late will have a 50% score reduction. These penalties can be waived with prior approval.

Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete and http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal respectively.

Scheduled Topics/Activities

Week	Session	Date	Topic
1	1	Mon, Aug 26	Intro to course
	2	Wed, Aug 28	**In-class exercise** - Intro to R - 1
	3	Fri, Aug 30	**In-class exercise** - Intro to R - 2
2		Mon, Sep 2	**No class** - Labor Day
	4	Wed, Sep 4	Statistical modeling and hypothesis testing
	5	Fri, Sep 6	**In-class exercise** - Gene Set Enrichment Analysis - **HW 1 due**
3	6	Mon, Sep 9	**In-class exercise** - Permutation tests
	7	Wed, Sep 11	Multiple testing and false discovery rates
	8	Fri, Sep 13	**In-class exercise** - FDR in gene expression analysis - **HW 2 due**
4	9	Mon, Sep 16	Dimensionality reduction and visualization with PCA and t-SNE
	10	Wed, Sep 18	**In-class exercise** - Visualizing human population genetic data
	11	Fri, Sep 20	Distance-based clustering - **HW 3 due**
5	12	Mon, Sep 23	**In-class exercise** - Hierarchical clustering of human populations
	13	Wed, Sep 25	K-means and medoids clustering
	14	Fri, Sep 27	**In class exercise** - Clustering of electronic medical records - **HW 4 due**
6	15	Mon, Sep 30	Maximum likelihood and apostiori probability
	16	Wed, Oct 2	Mixture models for clustering
	17	Fri, Oct 4	**In class exercise** - Osmotic stress in gene knockouts - **HW 5 due**
7	18	Mon, Oct 7	Linear regression
	19	Wed, Oct 9	**In class exercise** - RNA and protein levels
	20	Fri, Oct 11	Generalized linear models - **HW 6 due**
8	21	Mon, Oct 14	Multiple regression
	22	Wed, Oct 16	**In class exercise** - Protein evolutionary rates
	23	Fri, Oct 18	Feature selection - **HW 7 due**
9	24	Mon, Oct 21	Penalized likelihood
	25	Wed, Oct 23	**In class exercise** - Polygenic risk scores
	26	Fri, Oct 25	**In class exercise** - GWAS follow-up research - **HW 8 due**
10	27	Mon, Oct 28	Logistic regression and classification
	28	Wed, Oct 30	Naïve Bayes
	29	Fri, Nov 1	**In class exercise** - Classifying cell types from gene expression - **HW 9 due**
11	30	Mon, Nov 4	Support Vector Machines
	31	Wed, Nov 6	**In class exercise** - Protein secondary structure prediction
	32	Fri, Nov 8	Random forests - **HW 10 due**
12		Mon, Nov 11	**No class** - Veterans Day
	33	Wed, Nov 13	**In class exercise** - Predicting protein-protein interactions
	34	Fri, Nov 15	Evaluating classifiers - **HW 10 due**
13	35	Mon, Nov 18	**In class exercise** - Detecting mutations
	36	Wed, Nov 20	Neural networks
	37	Fri, Nov 22	**In class exercise** - TensorFlow playground - **HW 11 due**
14	38	Mon, Nov 25	Intro to final project, choose papers
		Wed, Nov 27	**No class**
		Fri, Nov 29	**No class** - Thanksgiving break
15	39	Mon, Dec 2	Final project work session - **Paper summary due**
	40	Wed, Dec 4	Final project work session
	41	Fri, Dec 6	Final project work session - **Draft slides due**
16	42	Mon, Dec 9	**Final presentations 1**
	43	Wed, Dec 11	**Final presentations 2**
17		Mon, Dec 16	Final paper due
		Wed, Dec 18	**Last day to turn in late work**

Honors Credit

Students wishing to contract this course for Honors Credit should email me to set up an appointment to discuss the terms of the contract. Information on Honors Contracts can be found at https://www.honors.arizona.edu/honors-contracts.

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Students are asked to refrain from disruptive conversations with people sitting around them during lecture. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Absence and Class Participation Policy

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at: http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop

The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable, http://policy.arizona.edu/human-resources/religious-accommodation-policy.

Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: https://deanofstudents.arizona.edu/absences

Participating in the course and attending lectures and other course events are vital to the learning process. As such, attendance is required at all lectures and discussion section meetings. Absences may affect a student's final course grade. If you anticipate being absent, are unexpectedly absent, or are unable to participate in class online activities, please contact the instructor as soon as possible. To request a disability-related accommodation to this attendance policy, please contact the Disability Resource Center at (520) 621-3268 or drc-info@email.arizona.edu. If you are experiencing unexpected barriers to your success in your courses, the Dean of Students Office is a central support resource for all students and may be helpful. The Dean of Students Office is located in the Robert L. Nugent Building, room 100, or call 520-621-7057.

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students.

Accessibility and Accommodations

At the University of Arizona we strive to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability or pregnancy, you are welcome to let me know so that we can discuss options. You are also encouraged to contact Disability Resources (520-621-3268) to explore reasonable accommodation.

Please be aware that the accessible table and chairs in this room should remain available for students who find that standard classroom seating is not usable.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as

described in the UA General Catalog. See: http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity.

The University Libraries have some excellent tips for avoiding plagiarism, available at http://new.library.arizona.edu/research/citing/plagiarism.

Selling class notes and/or other course materials to other students or to a third party for resale is not permitted without the instructor's express written consent. Violations to this and other course rules are subject to the Code of Academic Integrity and may result in course sanctions. Additionally, students who use D2L or UA e-mail to sell or buy these copyrighted materials are subject to Code of Conduct Violations for misuse of student e-mail addresses. This conduct may also constitute copyright infringement.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy

Our classroom is a place where everyone is encouraged to express well-formed opinions and their reasons for those opinions. We also want to create a tolerant and open environment where such opinions can be expressed without resorting to bullying or discrimination of others.

Additional Resources for Students

UA Academic policies and procedures are available at http://catalog.arizona.edu/policies

Student Assistance and Advocacy information is available at http://deanofstudents.arizona.edu/student-assistance/students/student-assistance

Confidentiality of Student Records

http://www.registrar.arizona.edu/personal-information/family-educational-rights-and-privacy-act-1974-ferpa?topic=ferpa

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.