

02-680: Essential Mathematics and Statistics for Scientists

Phillip Compeau and Seyoung Kim

Fall 2019

1. Course Information

1.1 Vital information

Course	Time	Lecture	
	Location	TR 3:00-4:20 PM	GHC 4307
Instructors	E-mail	Prof. Seyoung Kim sssykim@cs.cmu.edu	Prof. Phillip Compeau pcompeau@cs.cmu.edu
	Website	cs.cmu.edu/~sssykim/	compeau.cbd.cmu.edu
	Office	GHC 7721	GHC 7403
	Office Hour	Th 4:30-5:30 PM	W 10:30-11:30 AM
TAs	E-mail	Jun Ho Yoon junhoy@andrew.cmu.edu	Mihir Mongia mmongia@andrew.cmu.edu
	Office Hour	Tu 1:30-2:30 PM	M 3:00-4:00 PM
	Location	GHC 7607	GHC 7607

1.2 Course description

This course rigorously introduces fundamental topics in mathematics and statistics to first-year master's students as preparation for more advanced computational coursework. Topics are sampled from information theory, graph theory, proof techniques, phylogenetics, combinatorics, set theory, linear algebra, neural networks, probability distributions and densities, multivariate probability distributions, maximum likelihood estimation, statistical inference, hypothesis testing, Bayesian inference, and stochastic processes.

Students completing this course will obtain a broad skillset of mathematical techniques and statistical inference as well as a deep understanding of mathematical proof. They will have the quantitative foundation to immediately step into an introductory master's level machine learning or automation course. This background will also serve students well in advanced courses that apply concepts in machine learning to scientific datasets, such as 02-710 (Computational Genomics) or 02-750 (Automation of Biological Research). The course grade will be computed as the result of homework assignments, midterm tests, and class participation.

1.3 Pre-requisites

There are no formal prerequisites. However, we expect that students will have a strong foundation in high school mathematics (including calculus) and possess strong quantitative reasoning skills, as the course will be taught at a high level and proceed quickly.

1.4 Course Details

Canvas Homepage. The course homepage will be hosted on Canvas. Canvas will be used for attendance and as a central repository for grades. You should be automatically enrolled at <https://canvas.cmu.edu/courses/9677>.

Discussion Forum. An online forum is provided on Piazza as an area for discussion and questions. The forum will be moderated by the course staff who will respond to questions, but students are encouraged to help each other via discussion. However, assignment specifics should not be discussed — any hints will be provided by the teaching staff. You can find the class on Piazza at <https://piazza.com/class/jswou69xvn76qm>.

1.5 Textbooks

There are no required textbooks for this course. Below is an optional textbook on statistics if you would like to read further about the material covered in the statistics half of the course.

All of Statistics: A Concise Course in Statistical Inference by Larry Wasserman. This book is available in the CMU Engineering & Science library as a hardcopy and online resource.

2. Curriculum

2.1 Tentative course schedule

The lecture-by-lecture below is tentative and subject to change.

C O U R S E S C H E D U L E

Date	Lecture Topic	Date	Lecture Topic
8/27	Prologue and an overview of genome assembly	8/29	Proof by construction (Part 1): Euler's Theorem and Intro to Perfect Phylogeny
9/3	Proof by construction (Part 2): Perfect Phylogeny	9/5	Proof by contradiction: two proofs from the Greeks that every human should know
9/10	McCullough-Pitts neurons, perceptrons, and foundations of logic	9/12	Contrapositive, proof by cases, and the universality of binary neural networks
9/17	The Universality of NAND and Modularity in Biological Systems	9/19	Trees and Combinatorial Proofs: Induction and Double-Counting
9/24	Biological motif finding and information theory	9/26	Linear algebra 1: introduction to linear transformations
10/1	Linear algebra 2: The Invertible Matrix Theorem	10/3	Linear algebra 3: Eigenvalues and singular value decomposition
10/8	Real analysis 1: limits and continuity	10/10	Real analysis 2: universality of neural networks
10/15	Midterm 1	10/17	Probability, conditional probability, probabilistic independence
10/22	Random variables, pdf, cdf	10/24	discrete random variables
10/29	Continuous random variables	10/31	Expectation
11/5	Maximum likelihood estimation I: parametric inference	11/7	Maximum likelihood estimation II: sufficient statistic, exponential families
11/12	Multivariate distribution	11/14	Hypothesis testing and p -values I: confidence interval, Wald test
11/19	Hypothesis testing and p -values II: chi-squared test, t-test, and other tests	11/21	Permutation test, multiple testing
11/26	Bayesian inference		T H A N K S G I V I N G
12/3	Linear regression	12/5	Markov chain

3. Coursework

Coursework will consist of the following components. **No late assignments will be accepted.**

Homework assignments. (40% of grade) Written homework assignments will test your knowledge of the material covered in class.

Attendance and participation (10% of grade) Attendance will be taken, and we will have occasional in-class exercises that serve to reinforce the concepts we have covered. These exercises will not be graded, but participation will be expected in order to receive a complete grade for that day. You are allowed three “dropped” attendance grades without penalty. These can be used for any purpose.

Examinations. (50% of grade) The two will test your knowledge of the material from the class. The first midterm will be held in class, and the second midterm will be held during the university’s scheduled time. The exam dates are:

- Midterm 1 (25% of total grade): 10/15
- Midterm 2 (25% of total grade): Time and location TBD (will be posted when set by university)

The exams will not be cumulative: midterm 2 will cover material encountered after midterm 1. That having been said, later material in the class may build upon the techniques covered in the earlier material.

4. Collaboration Policy and Academic Integrity

All class work should be done independently unless explicitly indicated on the assignment hand-out. You may *discuss* homework problems with classmates, but must write your solution by yourself. If you do discuss assignments with other classmates, you must supply their names at the top of your homework. No excuses will be accepted for copying others’ work, and violations will be dealt with harshly. (Getting a bad grade is much preferable to cheating.)

The university’s policy on academic integrity can be found at the following link: <http://www.cmu.edu/academic-integrity/>. In part, it reads, “Unauthorized assistance refers to the use of sources of support that have not been specifically authorized in this policy statement or by the course instructor(s) in the completion of academic work to be graded. Such sources of support may include but are not limited to advice or help provided by another individual, published or unpublished written sources, and electronic sources.” You should be familiar with the policy in its entirety. **The default penalty for any academic integrity violation is failure of the course.**

In particular: use of a previous semester’s answer keys or online solutions for graded work is absolutely forbidden. Any use of such material will be dealt with as an academic integrity violation.

5. Other policies

Classroom etiquette: To minimize disruptions and in consideration of your classmates, we ask that you please arrive on time and do not leave early. If you must do either, please do so quietly.

The use of phones or other electronic devices during class is forbidden and will result in a zero discussion grade for the day (counts as missed class).

Excused absences: Students claiming an excused absence for an in-class exam must supply documentation (such as a doctor's note) justifying the absence. Absences for religious observances must be submitted by email to the instructor during the first two weeks of the semester. Note that job or internship interviews are not a justification for an excused absence.

Other: The following policies of 15-110 also apply to this class. This text is mostly quoted from the 15-110 website (with some modifications):

- **I must be out of town for a university related event (e.g. member of a team). What should I do about my assignments?**

If you have an official excuse we will make special arrangements for you to submit the assignment, please contact the instructors.

- **I am out of town attending a family/important event. How can I submit my assignments due for the week?**

The assignment must be submitted online before the due date.

- **I missed the in-class exam because I fell sick. What should I do?** You must immediately seek medical treatment and receive an official medical excuse. You must also contact the instructors prior to the exam or as soon as possible. If you can produce documentation we can make arrangements to give you a makeup test. Otherwise, we will be unable to make any exceptions.

- **I am failing the course. Is there any extra work I can do to get a passing grade?** Unfortunately, we cannot make exceptions. The best way to avoid this situation is to talk to one of the instructors as soon as possible to find out what you need to do. Do not wait until the last few weeks of classes to discuss your performance.

6. Accommodations for Students with Disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with us as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, we encourage you to contact them at access@andrew.cmu.edu.

7. Provost's Statement on Student Well-Being

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological

Services (CaPS) is here to help: call 412-268-2922 and visit their website at <http://www.cmu.edu/counseling/>. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

CaPS: 412-268-2922

Re:solve Crisis Network: 888-796-8226

If the situation is life threatening, call the police:

On campus: CMU Police: 412-268-2323

Off campus: 911

If you have questions about this or your coursework, please let us know.