
COGS 109

Introduction to Data Analysis & Modeling

Instructor Jacob Olson

Welcome to COGS 109!

As electronic data storage and computational abilities expand, skills in data analysis and modeling are becoming increasingly desired in both business and science. This course introduces you to the field of data science by presenting a variety of common big data analysis and modeling techniques. To accomplish this, each lecture will focus on one analysis technique. Lecture will be highly interactive with class time consisting of a mixture of group work, iClicker polling, large class discussion, and lectures. Techniques covered will include k-means clustering, Bayesian inference, principal component analysis (PCA), linear regression, perceptrons, and neural networks. You will use MATLAB to implement and utilize these analysis and modeling techniques through lab sections, take-home assignments, and a group project. Overall, this course will emphasize fundamental understanding of the purpose of analyses and will provide ample hands-on experience. This is done to produce individuals prepared in two key ways:

- for interaction with data science in everyday life and work.
- for continuing in-depth computational methods training.

Learning Objectives

By the end of this course you will be able to:

- interact confidently with data analysis and modeling techniques.
- utilize data science terminology correctly and be able to translate it into plain English.
- select and defend a given technique based on the dataset and objective.
- apply analysis techniques using MATLAB.
- visualize, interpret, and communicate results of analyses.

Resources

There are no reading materials to be purchased. We will provide all necessary materials online at the ThisCourse website. You will be asked to read and occasionally watch video tutorials online before attending class or for further learning afterwards.

UCSD Summer Session II 2015

Lecture

T Th CSB 002 11:00am - 1:50pm

Sections

A01 T Th CSB 115 9:00-9:50am

A02 T Th CSB 115 10:00-10:50am

Course Websites

For all course logistics, content, and announcements:

thiscourse.com/ucsd/cogs109/su15/

For class discussions, Piazza:

<https://piazza.com/ucsd/summer2015/cogs109/home>

For grades, TED: <http://ted.ucsd.edu/>

Teaching Team

Jacob Olson

Instructor

Email

j3olson@ucsd.edu

Office

Mandeville Coffee Cart

Office Hours

MW 1-2pm or by appointment

Melissa Troyer

Teaching Assistant

Email

mtroyer@ucsd.edu

Office

CSB 115

Office Hours

MW at 9-10am or by appointment

Materials Needed

iClickers will be used in class to enhance engagement on a daily basis. Clickers can be purchased at the bookstore. They must be registered for the class through the TED website.

Grading Policy

Item	Total Grade % (I will use the column that gives you the highest grade.)		Details
Class Participation	5% (3% extra credit also possible)	5% (3% extra credit also possible)	iClickers will be used in weeks 1-4 to enhance classroom discussions and assess the group comprehension of topics. Students will receive 1% total grade credit each day 75% of questions are answered. (See extra credit section for details)
Technique Assignments	20%	20%	Assignments will be given nearly every lecture. In class time will be structured towards completing portions of the assignments with support (so be there!). The remainder will be completed outside of class.
Midterm 1	20%	10%	Advice: Midterm exams are based strongly on the technique assignments. If you fully understand the homework, you will do well on the tests. Work hard to understand the concepts and complete your own code (academic integrity) on the assignments! Practice makes perfect!!
Midterm 2	20%	10%	
Class Project	15%	15%	Groups of 3-4 will select and complete an analysis project including presentation. Details to follow.
Final	20%	40%	Cumulative: Midterms + Project Skills

Standard percentage grading scale will be used, adjusted at my discretion for final letter grades. +/- will be used.

Extra Credit

Extra credit is available by completing the class participation for more than five lectures (up to 3%). Up to two SONA experimentation credits can be completed for 2% credit total, <https://ucsd.sona-systems.com>, or a two page research paper can be completed as an alternative. Contact the instructor for a topic. It is due Friday of Week 5.

Late Policy

No late assignments will be excepted unless there are extenuating circumstances validated beforehand by the instructor or with valid documentation (i.e. note from doctor). Please turn in what you have at the due date.

Regrade Policy

If you feel something has been misgraded, please write down what you feel is incorrect and resubmit to the instructor. Resubmission must be received within **one week** of the assignment being returned to the class. Please be warned that errors can be corrected either for or against you on a regrade, so be certain there has been an error. It is highly recommended to first discuss your results with another student or someone from the teaching team. This is a great way to learn from any mistakes you may have made.

Class Schedule*

Week	Tuesday	Thursday
1	Intro/Matlab Review	Clustering
2	PCA	Bayesian Inference
3	Midterm 1 / Linear Regression	Linear Regression / Neural Networks
4	Neural Networks	Midterm 2 / Project Workday 1
5	Workday 2 / Reinforcement Learning	Project Presentations

Final: Sat September 5th, 11:30-2:30

*All topics/midterms are subject to change. Notice will be given of a midterm date change at least one week in advance.

Learning Environment Considerations

Please be considerate to your fellow classmates and refrain from taking part in activities that may disrupt or distract from the class. Examples of such behavior are web browsing, chatting with neighbors, etc. The goal is to help everyone learn to the best of their ability.

Contact and Correspondence

I am more than happy to meet up with students and discuss the class and related material either during my office hours or in another prearranged time. Please email me or catch me before/after class to schedule a time if office hours do not work. It is my experience that email is not conducive for communicating or clarifying material, so please do NOT email with content questions. Those go to Piazza! I do not often check email during the day, so please expect up to 24 hours for a response when suggesting meeting times. (Ask me after class instead!)

Piazza Discussion Board

This term we will be using the Piazza website for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza. Please assist your fellow classmates, but be sure to do so in a manner that is consistent with the academic integrity policy (discuss the issue, but do not write the code needed or the answer). In addition to encouraging peer teaching and learning, engagement on Piazza can be a deciding factor for borderline grading cases (e.g., B+ vs. A-).

Find our class piazza page at: <https://piazza.com/ucsd/summer2015/cogs109/home>.

Office for Student Disabilities

Students requesting accommodations for this course due to a disability must provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD), which is located in University Center 202 behind Center Hall. Students are required to present their AFA letters to Faculty (please make arrangements to contact me privately) and to the OSD Liaison in the department in advance so that accommodations may be arranged. Please contact the OSD for further information: 858.534.4382 (phone) osd@ucsd.edu(email) <http://disabilities.ucsd.edu>(website).

Academic Integrity

I expect all students to honor the highest standards of academic integrity in this course. Please read and be familiar with the UCSD Policy on Integrity of Scholarship (<http://senate.ucsd.edu/Operating-Procedures/Senate-Manual/Appendices/2>). Overall, this means that all academic work is done by the individual who takes credit for the work, without unauthorized aid of any kind. For homework, this means you can work together, but you must write your own code without looking at that of others. For example: you could ask someone for help and discuss the code needed to accomplish your goal. You cannot, however, take notes of theirs and then immediately write the same thing. You must sit down and write your own code on your own, so that you know that you have processed what you have learned and can produce the necessary code **on your own**.

If you have any questions about what types of interactions are or are not allowed, please discuss them with someone from the teaching team **before** taking the actions.