

STA 141C - BIG DATA & HIGH PERFORMANCE STATISTICAL COMPUTING

Winter Quarter 2022

Instructor:	Bo Y.-C. Ning	Time:	TTh 9:00 – 10:20 am
Email:	bycning@ucdavis.edu	Place:	Young Hall 194/Zoom

TAs:

1. Wei Du wedu@ucdavis.edu
2. Eunseong Bae esbae@ucdavis.edu

Course Pages:

1. Canvas: submitting homework, grades, and lecture recordings
2. Piazza: lecture notes, homework, discussions [[Link to Piazza](#)](passcode: sta141c)

Class meeting times:

- **Lectures:** TTH 09:00-10:20 am Young Hall 194 (online until Jan 28, 2022)
- **Discussion:**
 - Sec. A01: W: 10:00 - 10:50 AM Olson Hall 147 (online until Jan 28, 2022)
 - Sec. A02: W: 11:00 - 11:50 AM Olson Hall 205 (online until Jan 28, 2022)

Survey: Beginning of the quarter [[Link](#)]

Office Hours: Tuesday 3:00 – 5:00 pm [[Zoom](#)] (Meeting ID: 965 3470 0118, Passcode: sta141coh)

TAs Office Hours:

- Wei: Wednesday 2:00 pm – 4:00 pm [[Zoom](#)] (no passcode)
- Eunseong: Friday 10:00 am – 12:00 pm [[Zoom](#)] (no password)

Objectives:

- Learn numeric linear algebra methods
- Learn popular algorithms for statistical models for big data
- Learn how to write python/R code efficiently for statistical data analysis
- Learn how to parallelize code for fast computation

Prerequisites: STA 141B (or STA 141A+ECS 010) (some linear algebra background is preferred)

Main References: This is a restricted list of various interesting and useful books that will be touched during the course. You are not required to purchase them but you may want to consult them occasionally.

- Gene. H. Golub and Charles F. Van Loan, 1996. *Matrix computations* (3rd/4th Edition), The Johns Hopkins University Press
- Kenneth Lange, 2010. *Numerical Analysis for Statisticians*, Springer

Tentative Course Outline:

I. Background (~ 1 week)

- Review of linear algebra

II. Numerical linear algebra; matrix computation methods (~ 4 weeks)

- GE; LU; Cholesky; QR; sweep operation; iterative methods

III. Optimization (~ 4 weeks)

- Newton's method; gradient descent; stochastic gradient descent; EM; MCMC; Variational inference

Grading Policy:

- **Participation** (8%) You must participate office hours (either TA's or mine) at least once and piazza at least once. Attendance does not count as participation.
- In order to make the participation grade transparent, you will log your in-person participation through a Google Form [\[Link\]](#). Each time you participate on Piazza, use the form to record it (within 1 week of when you participated).
- **Assignments** (60%) There will be 4 assignments. Short assignments will be graded for correctness. Long assignments will be graded with a rubric that measures the quality of your writing, graphics, and code. The rubric will be available on Canvas right after the first assignment.
 - Assignment grades are never dropped. If you can't turn in an assignment on time and have a legitimate excuse (such as medical emergency), email me before the deadline to arrange an extension or alternative.
- **Final Project** (20%) For the final project, you will work in teams of 3-4 people. Groups of 4 will be held to a higher standard than groups of 3. You need to find team members by yourself before the fourth week and submit a proposal before the sixth week. The project will be due in finals week. More details will be given after the third week.
- **Final Presentation** (10%) In week 10, each group will present preliminary results from their project to the class. Each group can either elect a leader to present results or choose to let everyone in the team to present. You are required attend and provide feedback on presentations from some of the other teams. More details about presentations will be released in week 8.
- **Extra credits:** There are two ways to receive bonus points
 - Homework (2 points per homework): receive full credit + show great effort (e.g., explain your code in great details or doing more than asked). TAs and graders have the right to make the judgement, but only less than 5% will receive this credit per each homework.

- Active participation: Since participation is the best way to get help and verify that you understand the material, active participants (e.g., ask good questions on Piazza or during office hours; actively answer other students' questions on Piazza) will be more likely to get an $A+$ or, e.g., move the grade from $C+$ to $B-$.
- **Grade scale:** $A-$ to students who score 90%, $B-$ to students who score 80%, $C-$ to students who score 70%, $D-$ to students who score 60%.
- We must be fair to everyone in the class and cannot raise a single student's grade, thus, please avoid sending emails to the instructor or to the TAs asking to raise one's score.

Tentative schedule (subject to change):

Homework #1	2nd week
Homework #2	4th week
Final Project Team member list due	4th week
Homework #3	6th week
Final project proposal due	6th week
Homework #4	8th week
Final Presentation	10th week
Final Report & Peer evaluation due . .	Final exam week

Note: No late homework will be accepted. If having difficulty for submitting homework on time due to special circumstances, you must send me a formal letter with formal proofs (e.g., medical report) **before** the deadline.

Piazza:

All questions regarding the course contents and organization, should be posted on Piazza so that all students can participate in the discussion. You are encouraged to answer course contents questions posted by other students.

NOTE:

- Please sign up for this class at: piazza.com/ucdavis/winter2022/sta141c (passcode: sta141c)
- Please do not post homework questions close to the homework submission's deadline: you may not be able to get the answer on time.
- The TAs and I are here to help you during office hours; if you want to ask questions outside the office hours, please post your question on Piazza. The TAs and I will monitor the channel, and answer when needed.
- Be polite and respectful to others.
- Search before you post. Your question may have already been asked and answered.
- When you post a question, explain the context and give an example of what you mean.

Code of conduct: Students must adhere to the UC Davis code of conduct <https://ossja.ucdavis.edu/code-academic-conduct> to an external site.. Violations of the code of conduct include (but are not limited to!) communicating and collaborating during midterms, copying, attempting to copy and letting someone copy a graded assignment; doing someone else's Homework/exam/project assignment; to have someone else doing

one's Homework/exam/project assignment; to share Homework/exam/project assignment; to submit work that is not yours. The fact that the violation did not benefit you directly, does not mean that it is less important.

One of the responsibilities delegated to faculty by the UC Davis Code of Academic Conduct is to report suspected academic misconduct to the Office of Student Support and Judicial Affairs (OSSJA). Thus, any violation will be reported, and students found guilty will get an F, no matter the extend and type of their violation. Please, do not do it.

Academic Honesty Professional programmers talk to their coworkers and use references to help solve programming problems, so I encourage you to:

- Discuss the problems with your classmates.
- Search for references online and in books.
- Adapt short pieces of code (≤ 10 lines) you find on Piazza or online. When you do this, you must cite the source. For Piazza, cite the post number. For other sources, cite the title, author, and URL.

That said, all writing and graphics must be your own work. At least 75% your code must be your own work. In addition, you must add comments to the code you wrote. If you're unsure whether something is okay, please ask!

Students need accommodations: Any student needs special accommodations (e.g. physical, learning, psychiatric, vision, hearing, etc.) who needs to arrange reasonable accommodations must contact the Student Disability Center (SDC). Faculty are authorized to provide only the accommodations requested by the SDC. If you have any questions, please contact the SDC at 530/752-3184 or sdc@ucdavis.edu.

Course materials: My lectures and course materials, including videos, lecture notes, discussions, tests, outlines, etc, are protected by U.S. copyright law and by University policy. I and the TAs are the exclusive owner of the copyright in those materials I create. You may take notes and make copies of course materials for your own use. You may also share those materials with another student who is enrolled in or auditing this course. You may not reproduce, distribute or display (post/upload) lecture notes or recordings or course materials in any other way — whether or not a fee is charged — without my express prior written consent. You also may not allow others to do so. If you do so, you may be subject to student conduct proceedings under the UC Davis Code of Academic Conduct. Similarly, you own the copyright in your original papers and exam essays. If I am interested in posting your answers or papers on the course web site, I will ask for your written permission.

Acknowledgement: The course material is developed based on the recommended textbook, on online resources as cited in the relevant materials, and on the courses taught by Dr. Cho-Jui Hsieh at the University of California, Davis (now at UCLA) and Dr. Hua Zhou at the University of California, Los Angeles.