STA 141C - Big Data & High Performance Statistical Computing

Spring Quarter 2022

Instructor:	Bo Y.-C. Ning	**Time:**	TTh 12:10 – 01:30 pm
Email:	bycning@ucdavis.edu	**Place:**	Wellman Hall 216

TA: Wei Du wedu@ucdavis.edu
Reader: Chenze Li czeli@ucdavis.edu
Course Pages:

1. Piazza. Lecture notes will be posted on Piazza. We will use piazza for posting and answering questions Please sign up Piazza [here] (password: sta141cSPR)
2. Canvas. Canvas is used for submitting homework, final report, and publishing grades

Class meeting times:

- **Lectures:** TTH 12:10 - 01:30 p.m. Wellman Hall 216
- **Discussion:**

  Sec. A01: W: 12:10 - 01:00 p.m. Wellman Hall 115

  Sec. A02: W: 01:10 - 02:00 p.m. Hoagland Hall 113

Survey: Beginning of the quarter [Link]

Office Hours: Tuesday 3:00 – 5:00 pm (Meeting ID: 949 4151 8117, Passcode: sta141coh)

TA's Office Hours:

- Wei: Thursday: 9:00 - 11:00 am (Meeting ID: 944 0011 5678) (no passcode)

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 ($\sim$ 1 week)
  - Review of linear algebra
- II. Numerical linear algebra; matrix computation methods ($\sim$ 4 weeks)
  - GE; LU; Cholesky; QR; sweep operation; iterative methods
- III. Optimization ($\sim$ 4 weeks)
  - Newton's method; gradient descent; stochastic gradient descent; EM; MCMC; Variational inference

Grading Policy:

- **Participation** (5%) You must participate on piazza at least once, either ask a question or answer a question. Attendance does not count as participation.

  To receive full credit, you need to fill out the google form: [link]
- **Assignments** (55%) There will be 4 assignments. 15% each for the first three assignments and 10% for the last assignment.
- **Final Project Proposal** (5%) At the end of the 6th week, each team need to submit a project proposal. The team can proceed with working on the proposed project if approved by the instructor.
- **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.
- **Final Project** (25%) For the final project, you will work in teams of 3-4 people. You need to find team members by yourself before the third 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.
- **Extra credits**: If you actively participate in the class and Piazza and/or do excellent on homework, I would like to consider giving extra credits for your final score if you are 1 or 2 points away from the next grade scale. Thus, it would be best if you made yourself visible in class.
- **Tentative 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: Homework submitted within 12 hours after the deadline will receive 80% of the final score. Homework submitted between 12 and 24 hours after the deadline will receive 50% of the final score. Submission later than 24 hours after the deadline will receive 0 points. Please send the email to our reader (TBA) directly regarding late submission.

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/class/l0rdq4ww7jh6vi (password: sta141cSPR)
- Please use the same name as it on Canvas.
- Please use ucdavis.edu to sign up Piazza.
- 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 TA 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 TA 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. For posts do not follow this rule, I will remove it.

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 ($\leq$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.