STA 208 - Statistical Machine Learning

Spring Quarter 2022

Instructor:	Bo Y.-C. Ning	**Time:**	TTh 9:00 – 10:20 AM
Email:	bycning@ucdavis.edu	**Place:**	Olson Hall 106

Course Pages:

1. Piazza. Lecture notes will be posted on Piazza. We will use piazza for posting and answering questions
   - Please sign up Piazza at: `https://piazza.com/class/l15ifr7n5h912m` (password: sta208)
2. Canvas. Canvas is used for submitting homework, final report, and publishing grades

TAs: We have two TAs:

- Lead TA: Xiner Zhou (`xezhou@ucdavis.edu`). Role: monitoring piazza, holding office hours, and leading discussion sessions
- TA: Jing Lyu (`jjlyu@ucdavis.edu`). Role: grading homework. Any questions regarding homework grade, please contact her directly.

Office Hours (on Zoom):

- Instructor: every Monday, 2:00pm - 4:00pm [Zoom link] (password: sta208oh)
- Lead TA: every Thursday, 3:00pm - 5:00pm (Zoom: 984 5429 7603 password: 208)
- Discussion sessions are used as additional office hours

Waitlist: Unfortunately, we are not going to increase the class size. This class has a long waitlist. If you decide you don't want to take the class, please drop immediately to make room for others. If you are in statistics or biostatistics major that are required to take the course, please send me an email.

Main References:

- Prof. James Sharpnack posted his lectures on Youtube. Most materials covered in the lectures will be similar to his. It is always good to learn the same topics from two different perspectives. I encourage you to watch his videos here: [videos].
- There is no required textbook: all materials will be available on the class website.
- A solid primer is An Introduction to Statistical Learning, by James, Witten, Hastie, and Tibshrani. However, it takes a very different approach from us and only partially overlaps on material.
- A great advanced textbook is Elements of Statistical Learning by Hastie, Tibshrani, and Friedman, the book is free at `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`. We will only cover some of the material.

- Also good, but still advanced, are Pattern Recognition & Machine Learning by Bishop and (the encyclopedic) Machine Learning by Murphy. Both of these books introduce the material from a more computer engineering (or AI), rather than statistical, perspective.
- Other useful and more practically oriented books are Data Science for Business by Provost and Fawcett and Predictive Analytics by Siegel. Both of these books emphasize principles rather than technical details.
- Relevant papers will be provided in each lecture.

Objectives: Learn various methods in statistical machine learning; learn the reasoning for using those methods; learn to use python for analyzing large scale dataset with computational intensive state-of-the-art statistical methods

Prerequisites: STA 206; STA 207; STA 135; Or their equivalents.

Tentative Course Outline:

- Topic 1: Linear methods for regression (3 weeks)
  - Linear regression, bias-variance trade-off, subset selection, ridge regression, lasso, lasso in high-dimension
- Topic 2: Classification (2 weeks)
  - Linear method for classification, logistic regression, unsupervised learning, support vector machine, multiclass classification, K-means, nearest neighbor classifiers, mixture models, PCA, factor analysis, PCA in high-dimension
- Topic 3: Nonparametric methods (2 weeks)
  - Kernel density estimation, tree-based method, basis expansions
- Topic 4: Bayesian methods (2 weeks)
  - Intro to Bayesian methods, Bayesian nonparametric method - Gaussian processes, Bayesian high-dimensional method - sparse priors
- Topic 5: Other topics if time permits
  - Causal inference, graphical models, neural networks, variational inference, topological data analysis, manifold estimation

Grading Policy: Homework (65%), Final project (35%)
Grading scale: A- ≥ 90; B- ≥ 80; C- ≥ 70; D- ≥ 60; and fail if < 60.

Grading and important dates:

Warming up exercise	Due April 8; 5 points
Team member list	Due April 17
Homework #1	Due April 24; 15 points
Homework #2	Due May 6; 15 points
Final project proposal due	Due May 15; 5 points
Homework #3	Due May 20; 15 points
Homework #4	Due June 3; 15 points
Final Presentation	10th week; 10 points
Final Report	Due June 06; 20 points
Peer review	Due June 07 11:59am

Late submission: 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 the supporting TA directly about late submission. Late submission for the final report will not be accepted.

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.

Notes:

1. Please sign up for this class at: https://piazza.com/class/l15ifr7n5h912m (passcode: sta208)
2. Please do not post homework questions close to the homework submission's deadline: you may not be able to get the answer on time.
3. 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. Emails and private messages regarding lectures and homework will be ignored.
4. Be polite and respectful to others.
5. Search before you post. Your question may have already been asked and answered.
6. 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 ($\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. James Sharpnack at the University of California, Davis.