STA 035C: Statistical Data Science III

Syllabus (as of March 26, 2025) - Spring 2025

Basic Course Information

Instructor: Dogyoon Song

- *Email:* `dgsong@ucdavis.edu`
- *Office hours:* Wednesdays 4–5 PM, or by appointment (*Office:* MSB 4220)

Teaching Assistant: Soobin Kim

- *Email:* `sbbkim@ucdavis.edu`
- *Office Hours:* Mondays and Thursdays 1-2 PM (*Location:* TBA)

Time and Location: This is an in-person only class, and lectures will not be recorded.

- *Lectures:* Mondays, Wednesdays & Fridays, 12:10–1:00 PM (WELLMN 26)
- *Discussions (labs):* Run by TA
  - Section A01: Tuesdays, 10:00–10:50 AM (TLC 2212)
  - Section A02: Tuesdays, 11:00–11:50 AM (TLC 2212)

Course Description

This is the third and final course in the freshman data science sequence (STA 35A/B/C), offering a comprehensive introduction to statistical learning. Its primary goal is to help students understand fundamental methods, learn when and how to apply them, and recognize their limitations. Topics include a range of supervised and unsupervised techniques, cross-validation, the bootstrap, simultaneous inference, PCA, and nonparametric smoothing.

Prerequisites: The official prerequisite is as follows:

- STA 035B C- or better
- (MAT 016B C- or better or MAT 017B C- or better or MAT 021B C- or better)

Textbook: Please use the correct edition. Some of the HW problems will be the exercises from the textbook.

- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R (2nd ed.).* Springer Texts in Statistics. Springer, New York.
  (The book is available online here or here).

Online platforms: We will use multiple online platforms and resources in STA 035C.

- The course webpage will host lecture notes, homework, supplementary materials, etc.
- Canvas will be used for lab materials and for turning in homework. Solutions will be posted on Canvas.
- Piazza will be used for discussion (more details below).

Course Topics and Tentative Schedule

Topics: Below is a tentative list of topics that will be covered in this course.

- Basics in probability theory
- Introduction to supervised learning
  - Linear regression and some extensions
  - Classification, focused on logistic regression and Linear Discriminant Analysis (LDA)
  - Basics of nonparametric smoothing
- Model assessment, selection, and inference
  - Cross-validation and the bootstrap
  - Model selection and regularization
  - Simultaneous inference
- Unsupervised learning
  - Dimension reduction: principal component analysis (PCA)
  - Clustering: k-means and hierarchical

Course Schedule: Below is a tentative schedule, subject to changes depending on time constraints.

Week	Topics	Readings	Notes
1	Overview / Probability	JWHT, Ch 2.1	"HW 0"
2	Regression	JWHT, Ch 3	HW 1
3	Regression / Classification	JWHT, Ch 2.2 & Ch 4	HW 2
4	Classification	JWHT, Ch 4	Midterm 1
5	Resampling methods	JWHT, Ch 5	HW 3
6	Model selection	JWHT, Ch 6	HW 4
7	Simultaneous inference	JWHT, Ch 13	Midterm 2
8	Nonparametric smoothing	JWHT, Ch 7	HW 5
9	PCA / Clustering	JWHT, Ch 12 & Ch 6.3	HW 6
10	Clustering / Course recap	JWHT, Ch 12	Final exam

Assessments & Grading

Grading scheme: The students' performance in this course will be evaluated based on the following:

- **Homework** (30%)
- **Midterm exams** (30%)
- **Final exam** (40%)
- **Participation** (up to 3% extra)

Cutoffs for letter grades follow the standard UC Davis grading scheme:

- A: 90% or higher
- B: 80% or higher
- C: 70% or higher

- D: 60% or higher
- F: 59% or lower

These thresholds may be adjusted at the end of the quarter in a way that favors the students. For example, the cutoff for an "A" may end up being adjusted to be lower than 90%, but not higher.

Homework (30%): Homework assignments provide essential practice for students to internalize the statistical concepts and methodologies covered in class. It is essential for students to complete all of the homework assignments to deepen their understanding of the course material and prepare for exams. Typically, homework problem sets will be released on Wednesday morning, and will be due the next Tuesday at 11:59 PM PT. Please always check the course webpage and Gradescope/Canvas to ensure the correct due date. Before the course begins, students should complete the "Homework 0" for self-assessment; this will not be collected or graded, and no solutions will be provided.

- **Number of assignments.** There will be six homework assignments in total, not counting the preparatory "Homework 0" that you do not need to submit. You will have one week to complete each assignment. The lowest homework score will be dropped, so your final grade for the homework component is based on your five highest homework scores.
- **Submission requirements.**
  - All homework must be submitted as a single PDF file on Canvas by the due date. Email submissions are not accepted.
  - Please make sure to include "STA 35C," your name, and the last four digits of your student ID on the front page.
  - Homework may contain coding problems that require writing a combination of code and written prose. Your solutions to such questions should be prepared in R Markdown (or Quarto) and must be submitted in PDF format, specifically the result of calling "Knit PDF" from RStudio on your R Markdown document. All code used to produce your results must be shown in your PDF file (i.e., do not use 'echo = FALSE' or 'include = FALSE' as options anywhere). Rmd files do not need to be submitted, but may be requested by the TA and must be available when the assignment is submitted.
  - Non-coding portions may be typed (in LaTeX or any other software you are comfortable with), or written and scanned. The final submission must be a single PDF file containing all your coding- and non-coding- solutions to the assigned problem set.
- **Late submission policy.** ***No late homework will be accepted for any reason; submissions after the deadline will receive 0 points.*** However, your lowest homework grade will be dropped so that your grade for the 6 homework problem sets will be the average of your 5 highest homework scores. This dropped score is meant to accommodate emergencies or other extenuating circumstances, so no further arrangements or exceptions will be made.

  It is highly recommended that you begin working on each homework assignment as soon as it is released. Please avoid posting homework-related questions close to the submission deadline, as you may not receive a timely response.

Midterms (30%) and Final (40%): There will be two midterms and one final exam.

- *Midterm exam 1:* Friday, April 25, 2025, 12:10-1:00 PM (Wellman Hall 26)
- *Midterm exam 2:* Friday, May 16, 2025, 12:10-1:00 PM (Wellman Hall 26)
- *Final exam:* Friday, June 6, 2025, 1:00-3:00 PM (Wellman Hall 26)

The midterms will be held in class, during the scheduled class times. The final exam will take place at the designated time noted above. More details on exams will be announced at a later date.

There will be no make-up exams. The lower score of the two midterms will be dropped. If you must miss an exam due to illness, travel, or any other reasons, this exam will be the one to be dropped. For the final, if you have another final exam starting or ending 30 minutes before or after the scheduled time, you may present documentation and request to start 15 minutes earlier or later. To request this accommodation, the student must contact the instructor at least two weeks in advance for the request to be processed.

Participation in lectures (up to 3% extra): Students may earn up to three bonus points through active participation. Of the three total bonus points:

- **Up to two points** will be awarded at the instructor's discretion, based on overall participation and engagement in lectures. To qualify for ***at least one*** of these extra points, you must ***contribute to a class discussion in person during a lecture*** (not in discussion sessions) at least once any time during the quarter. Contributions include asking or answering a question, or providing a relevant comment on a presentation. Then, please respond to the survey (in the Quizzes tab on Canvas) indicating the date of the lecture you contributed and a short description of your question or comment.
- **Up to one** point will be awarded by the TA's recommendation, based on participation in discussion sessions and TA office hours.

Grade disputes and adjustments: Students have 24 hours after receiving a grade on any assignment to contest it. Grading will be consistent and we will provide detailed rubrics. If you believe you deserve a different grade, prepare a strong argument and submit it by email to the TA.

Course Policies

Attendance & Participation

Lectures: Although attendance is not mandatory, it is crucial for your success to attend each lecture and participate in discussions. ***Classes will be held in person and will not be recorded*** unless otherwise specified. If you miss a lecture, you are responsible for any material covered and announcements made in your absence; in that case, studying on your own and attending office hours is recommended if you have further questions.

Active participation is encouraged, both in class and on Piazza. Cell phones, laptops, and other electronic devices must be silenced in class. Laptops and tablets should be used in class only for learning purposes related to the lecture. Chatting or other disruptive behavior may result in a student being dismissed for the day; repeated disruptive behavior may result in a discussion with your academic advisor on appropriate further action.

Discussion sessions: It is also important to attend discussion sessions, where R tutorials and additional examples will be provided to deepen your understanding on the lecture content. These sessions are designed for students' practice through hands-on experience, and have no separate assessments during the discussion sessions. We value your ability to communicate mathematical and statistical ideas accurately, so please take the opportunity to share your thoughts and questions, and be prepared to present solutions to the group.

Code of conduct & Academic integrity

All students are expected to follow the UC Davis Code of Academic Conduct. Violations include (but are not limited to) collaborating or communicating during exams, copying or allowing someone to copy graded assignments, doing someone else's homework/exam (or having someone do yours), sharing assignments/exams, and submitting work that is not your own. The fact that the violation did not benefit you directly does not diminish its seriousness.

Under the UC Davis Code of Academic Conduct, faculty are required to report suspected academic misconduct to the Office of Student Support and Judicial Affairs (OSSJA). Any violation will be reported, and students

found guilty will receive an F, regardless of the extent or type of violation. A first offense leads to a failing grade on the relevant assignment or exam; a second instance results in a failing grade in the course. Any student who cheats on an assignment or exam will be referred to the Office of Student Support and Judicial Affairs. More information on academic dishonesty and UC Davis policy is available on the OSSJA website. Please do not violate the code of conduct—remain vigilant and avoid any misconduct.

Collaboration and the use of large language models: Collaboration in studying course material is encouraged, and students are welcome to discuss assignments with classmates. However, all submitted work must be your own. If you collaborate on homework, clearly list the names of all students involved.

Use of large language models (LLMs) such as ChatGPT, Grok, etc. is not permitted. Any suspected use will be reported to OSSJA.

Distribution of course materials

Please do not distribute any course materials outside of this class, as doing so results in an infringement of copyright as per UC policy. Use of sites like Course Hero and Chegg is not permitted. You may take notes, make copies of course materials for your own use, and share them with other students who are auditing this course. However, you may not reproduce, distribute, or display (post/upload) lecture notes or course materials in any other way—whether or not a fee is charged—without the instructor's express prior written consent. You also may not allow others to do so. Violations may result in student conduct proceedings under the UC Davis Code of Academic Conduct.

Changes to syllabus

This version of the syllabus is current as of March 26, 2025. The instructor may update it based on student progress or shifting instructional priorities. Any updates will be communicated promptly.

Getting Help

Piazza, office hours, and Email

Piazza You can access Piazza via the link on the Canvas sidebar, or directly through this link. Students are encouraged to answer each others' questions; the TA will moderate by checking in every day. The quickest way to get a question answered is likely on Piazza, since anyone in class can answer. Please follow the guidelines below when posting on Piazza:

- **Please be respectful.** Any content deemed inappropriate will be removed by the TA, and reported to the instructor.
- **Please search before posting.** Your question may have already been answered.
- **Please be specific and informative.** For example, posting a code and simply asking "what is wrong with this code?" is not acceptable. Please explain what you have tried and where you are stuck in detail. If appropriate, along with your posted question, explain what else you tried that didn't work. The answers to many common coding questions can be found on https://stackoverflow.com/.

  For answerers, "I don't know either" is not an appropriate answer as it does not contribute constructively to the conversation.

Office hours: Who to ask?

- Dogyoon Song (Instructor): Questions regarding lecture content or organizational aspects of the course.
- Soobin Kim (TA): Questions regarding contents of the course, discussions, homework assignments and their grading, code, Canvas and Piazza.

Email Please use email only for questions related to private matters (accommodations, grading, emergencies, etc.). For questions about class logistics or content, please post on Piazza, ask in class, or bring them to discussion sessions or office hours. If you must ask a question about a non-private matter via email, you must first document how you attempted to resolve it (e.g., "I double checked the syllabus," "Piazza answers were conflicting," ...). Emails not following these guidelines may not be answered.

Please do not send messages on Canvas; the Canvas inbox is not monitored and you may not receive a response.

Other campus resources

- **Statistics Tutors** at the Academic Assistance and Tutoring Centers provide support for RStudio. More information is available here.
- **UC Davis Student Resources** Many students face different challenges during college, and it is healthy to seek support. This is a comprehensive list of resources covering general academics, health and wellness, finances, housing, career/internship, and other topics.
- **Health & Wellness** Health and wellness resources are available here. If you have an emergency, call 911 immediately, or go to the nearest emergency room. Mental health staff are available 24 hours/7 days week by phone at (530) 752-2349 (follow the prompts to speak with a counselor).

Accommodations for students with disabilities

UC Davis is committed to educational equity in the academic setting, and in serving a diverse student body. All students who are interested in learning more about the Student Disability Center (SDC) are encouraged to contact them directly at `https://sdc.ucdavis.edu`, `sdc@ucdavis.edu` or (530) 752-3184. If you are a student who requires academic accommodations, please submit your SDC Letter of Accommodation as soon as possible, ideally within the first two weeks of class.