# SDS 391P 3-Theory of Linear Models

**Department of Statistics and Data Sciences**
**The University of Texas at Austin**
**Fall 2025**

**Instructor:**
Alessandro (Ale) Rinaldo, Professor, Department of Statistics and Data Sciences
Email: alessandro.rinaldo@austin.utexas.edu
Office: WEL 5.216G
Office Hour: by appointment.

**TA:**
Khai Nguyen, Ph.D. student, Department of Statistics and Data Sciences
Email: khainb@utexas.edu
Office Hour: TBA

**Lectures:**
Tuesday and Thursday, 9:00am - 10:30am, FAC 101B.

**Class Website:**
https://arinaldo.github.io/Teaching/SDS391-3_F25/SDS391-3_F25_index.html
Please check the website on a regular basis.
**Note:** homework submissions and solutions will be handled via Canvas.

**Prerequisites:**
Enrollment into the Ph.D. program in Statistics and Data Science or instructor's approval.

**Course Description:**
SDS 391-3 is an intermediate graduate course in theoretical statistics for PhD students, covering two separate but interrelated topics: (i) stochastic convergence, (ii) learning theory and (ii) linear regression modeling. The material and style of the course will skew towards the mathematical and theoretical aspects of common models and methods, designed to provide a foundation for those who wish to pursue research in statistical methods and theory. This is not an applied regression analysis course.

**Course Objectives:**
The objectives of the course are three-fold.

- to describe various modes of stochastic convergence, illustre their difference, exemplify their uses in various problems and provide tool-box of basic techniques for carrying out asymptotic analyses of statistical and probabilistic models.
- to provide a foundational understanding of linear modeling, including linear algebra background, the theoretical underpinnings of well-known estimators, tests, model diagnostics, methods for variable selection and dimension reduction and minimax lower bounds. Linear models will be introduced and studied mainly under an "assumption-lean" framework in which minimal assumptions are imposed on the data generating distribution – notably, the

regression function will not assumed to be linear and the model may very well be mispecified. The study of linear models will be the main component of the course.

- to learn basic theoretical and algorithmic concepts in learning theory, of which linear regression is a special case.

Successful completion of the course should give students a firm theoretical foundation to access the literature on a variety of modern statistical methods and theory.

**Target Audience:**
As this course is a required core course for the Ph.D. in Statistics and Data Sciences, the target audience is graduate students pursuing a Ph.D. in statistics or another closely related field (in which case instructor's approval is required). While there are no formal course prerequisites, students are expected to have graduate-level knowledge of probability, mathematical statistics, and linear algebra.

**Course component and policies:**

Lectures:. Lectures for this course will be held in person but, if/when needed broadcast via zoom. In the event that lectures are held via zoom, students are expected to attend lectures synchronously, but special arrangements can be made in coordination with the instructor if necessary.

Problem Sets and Homework: Biweekly homework assignments will be a combination of textbook-style exercises and problems taken from the literature. Students are encouraged to work collaboratively with each other on homework problems, but should make a full individual effort before consulting or comparing with peers. **There will be approximately 6 homework assignments**. Late homework submissions will not be accepted unless approved by the instructor. If you are unable to meet the homework deadline, please contact the instructor as soon as possible.
Homework Submission: Homework assignments are to be submitted electronically via Canvas. You must provide a clean, easily readable scan of your assignment, e.g., through a scanner or your phone or computer. tablet, ect. Unintelligible solutions will receive null grades. If you are writing out your answers by hand, please use a dark pencil to enhance the contrast and readability. When submitting multi-page assignments, you are responsible for numbering the pages and scanning them in the correct order. Assignments with pages out of order, missing page numbers, or without the student's name will not be graded and will automatically receive a 0 score.
Academic Integrity for Homework: You are responsible for ensuring that any answers you submit for evaluation are the result of your own efforts. Final solution writeups and any requested computer code must be a result of your own work, and no material may be copied from another student. If you work with other students on the homework problem sets, **you must turn in your own completed assignment** and please be sure to include the names of your collaborators with your assignment.

Final Project: The final project involves picking a topic of interest, reading the relevant results in the area and then writing a short review paper (8 pages max) summarizing the key ideas in the area. You may focus on a single paper if you prefer. Your are NOT expected to present your own, novel research, but you are welcome to. The paper should include background, statement of important results, and brief proof outlines for the results. If appropriate, you should also include numerical experiments are an application with real data.

Information and guidelines.

- You may work by yourself or in teams of two.
- The goals are (i) to summarize key results in literature on a particular topic and (ii) present a summary of the theoretical analysis (results and proof sketch) of the methods (iii) implement some of the main methods. You may develop new theory if you like but it is not required.
- You will provide: (i) a proposal, (ii) a progress report and (iii) and final report.
- The reports should be well-written.

Timeline.

- Proposal. Due October 3. A one-page proposal. It should contain the following information: (1) project title, (2) team members, (3) precise description of the problem you are studying, (4) anticipated scope of the project, and (5) reading list. (Papers you will need to read).
- Progress Report. Due November 7. Three pages. It should include: (i) a high quality introduction, (ii) what have you done so far, (iii) what remains to be done and (iv) a clear description of the division of work among teammates, if applicable.
- Final Report. Due December 12. The paper should be in NeurIPS format. (pdf only). No appendix is allowed. You should submit a pdf file electronically. It should have the following format:
    1. Introduction. Motivation and a quick summary of the area.
    2. Notation and Assumptions.
    3. Key Results. Proof outlines for the results.
    4. Implementation (simulations or real data example.)
    5. Conclusion and/or future work/open questions (if appropriate).

Class Recordings: In the event that course lectures are held via zoom and recorded, class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

**Course Grading:**
Your assessment and grades will be determined as follows:

- Homework assignments (70%).
- Attendance and class participation (10%).
- Final Project (20%).

**Course Material:**
The lectures will be based on material extracted from various sources.

**Stochastic convergence**

- T.S. Ferguson (1996). A Course in Large Sample Theory, Chapman.
- A. W. van der Vaart (1998). Asymptotic Statistiics, Cambridge University Press

- The notes used in my old Advanced Probability Overview course, last taught at CMU in 2020.
- E. L. Lehmann and J. P. Romano (2002). Testing Statistical Hypotheses, 4th edition, Springer.

Other recommended references:

- D. Hunter, Notes for a graduate-level course in asymptotics for statisticians, available here
- R. Serfling (1980). Approximation Theorems of Mathematical Statistics, John Wiley, New York.
- J. Shao(2003). Mathematical Statistics, 2nd edition, Springer.
- A. DasGupta (2008). Asymptotic Theory of Statistics and Probability, Springer.

**Learning Theory**

- Learning Theory from First Principles, by F. Bach (2025), available here.

**Linear modeling and linear algebra**

- R. Christensen (2020). Plane Answers to Complex Questions: The Theory of Linear Models, 5th Edition, Springer.
- G. A. F. Seber and A. J. Lee (2003). Linear Regression Analysis, Second Edition, Wiley.
- S. Weisberg (2013) Applied Linear Regression, 4th edition, Wiley.
- W. Greene (2017)., Econometric Analysis, 8th edition, Pearson.
- R. Christensen (2019). Advanced Linear Modeling: Statistical Learning and Dependent Data, 3rd edition, Springer.

**Minimax theory**

- A. Tsybajov, (2009). Introduction to Nonparametric Estimation. Springer.
- M. Wainwright (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint, Cambridge University Press.
- Y. Polyanskiy and Y. Wu () Information Theory: From Coding to Learning, Cambridge University Press, forthcoming available here.

**Sharing of Course Materials is Prohibited:**
No materials used in this class, including, but not limited to, lecture hand-outs, videos, assessments (quizzes, exams, papers, projects, homework assignments), in-class materials, review sheets, and additional problem sets, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well

aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course. In particular, use of any assignments provided through previous offerings of the course (e.g., previous semesters' homework or exams) or communication about details of these materials with students who have taken the course previously will be regraded as academic dishonesty in this course.

**Academic Integrity Expectations:**

Students who violate University rules on academic misconduct are subject to the student conduct process. A student found responsible for academic misconduct may be assigned both a status sanction and a grade impact for the course. The grade impact could range from a zero on the assignment in question up to a failing grade in the course. A status sanction can include a written warning, probation, deferred suspension or dismissal from the University. To learn more about academic integrity standards, tips for avoiding a potential academic misconduct violation, and the overall conduct process, please visit the Student Conduct and Academic Integrity website at http://deanofstudents.utexas.edu/conduct.

**Accessible, Inclusive, and Compliant Statement:**

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). Please refer to D&As website for contact and more information: http://disability.utexas.edu/. If you are already registered with D&A , please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.