MA 589-2B STATISTICAL TECHNIQUES FOR MACHINE LEARNING AND BIG DATA SYLLABUS

Semester: Fall 2025
Course Instructor: Keren Li, PhD
Contact Information: kli@uab.edu
Office: University Hall 4041
Office Hours: Tue 11:00 am - 12:00 pm

Course Info

Meeting times: Lectures: Tue/Thu 9:30 am - 10:45 am
Meeting location: University Hall 4002
Prerequisite: Grade of B or better in MA 585 or grade of B or better in MA 586 or equivalent.
Credits: 3 semester hours
Recommended Textbooks:

An Introduction to Statistical Learning: with Applications in R, 2nd Edition, Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 2021.
The Elements of Statistical Learning, 2nd Edition, Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome, 2009.
Computer Age Statistical Inference: Algorithms, Evidence, and Data Science, Efron, Bradley, and Trevor Hastie, 2016.

Course Description

This course provides a principled introduction to Statistical Machine Learning, with a focus on understanding both the statistical assumptions and the computational foundations behind modern learning algorithms. We begin by framing learning problems through the lens of optimization and model-based inference, then study core techniques including linear regression, classification, model selection, support vector machines, tree-based methods, and unsupervised learning.

Emphasis is placed on interpretable models, algorithmic reasoning, and the bias-variance tradeoff. Students will use the R programming language to implement methods, analyze data, and complete a hands-on final project that applies learned techniques to real-world datasets.

Learning Outcomes

Upon successful completion of the course, a student can

(1) **Comprehend Fundamental Concepts:** Develop a strong understanding of fundamental concepts in statistical machine learning, including supervised and unsupervised learning, model evaluation, regularization, and dimensionality reduction.
(2) **Apply Techniques:** Gain hands-on experience by applying various machine learning techniques, such as linear regression, logistic regression, decision trees, support vector machines, clustering, and more, to real-world datasets.

(3) **Program in R:** Acquire proficiency in the R programming language to implement and experiment with machine learning algorithms, analyze data, and visualize results.
(4) **Interpret Results:** Learn to interpret the outcomes of machine learning models, understand the implications of different evaluation metrics, and make informed decisions based on model performance.
(5) **Critical Thinking:** Cultivate critical thinking skills by examining the trade-offs between model complexity and generalization, understanding bias-variance trade-offs, and exploring the practical challenges of applying machine learning techniques.
(6) **Real-world Applications:** Recognize the diverse applications of machine learning in various fields, such as biology, finance, healthcare, and more, and gain the ability to adapt and apply learned techniques to address real-world problems.
(7) **Collaboration and Communication:** Enhance teamwork and communication skills through group discussions, collaborative projects, and presentations, fostering effective communication of technical concepts.
(8) **Ethical Considerations:** Understand the ethical implications and potential biases in machine learning models, and develop a sense of responsibility in applying these tools in a fair and unbiased manner.
(9) **Foundation for Further Learning:** Establish a solid foundation in statistical machine learning that prepares students for more advanced courses, research opportunities, and professional roles in data science, artificial intelligence, and related fields.

Course Content

Introduction to Statistical Learning: Overview of machine learning and its applications. The distinction between supervised, unsupervised, and semi-supervised learning. Model evaluation and generalization. Introduction to the unified view of learning as optimization and emphasis on model assumptions.

Linear Methods for Regression: Simple and multiple linear regression. Model fitting, parameter estimation, and hypothesis testing. Extensions to polynomial regression and regularization.

Model Assessment and Selection: Bias-variance trade-off and overfitting. Regularization methods (L_1, L_2) for improving model robustness. Cross-validation and its application in model selection. Performance metrics for classification and regression tasks.

Linear Methods for Classification: Logistic regression for binary classification. Multiclass classification using one-vs-rest and one-vs-one approaches. Maximum likelihood estimation and decision boundaries.

Basis Expansion and Kernel Methods: Feature transformation and basis expansion techniques. Kernel methods for nonlinear classification and regression.

Support Vector Machines: Maximum margin classifiers and support vector classification. Nonlinear separation using kernel trick. Soft-margin classification and handling outliers. Introduction to Stochastic Gradient Descent (SGD), Backpropagation, and the Energy-Based View of learning.

Tree-based Methods and Boosting: Decision tree construction and split criteria. Ensemble learning with bagging and boosting. Introduction to random forests and gradient boosting.

Unsupervised Learning: Clustering algorithms (K-Means, Hierarchical clustering). Density estimation with Gaussian Mixture Models (GMM). Principal Component Analysis (PCA) for dimensionality reduction.

Optional Advanced Topics: To be included based on course pacing and student interest: Autoencoders (AEs), Restricted Boltzmann Machines (RBMs), Energy-Based GANs, Neural Network Blocks and Attention Mechanisms, Meta-learning, and Reinforcement Learning.

Course Grade

Your grade in the course is determined by the points earned throughout the semester.

Homework Assignments: (50% total):

Preparation: R Basics and Tools (8%)
Chapter 2: Linear Methods for Regression (7%)
Chapter 3: Model Assessment and Selection (7%)
Chapter 4: Linear Methods for Classification (7%)
Chapter 5: Basis Expansion and Kernel Methods (7%)
Chapter 6: Support Vector Machines (7%)
Chapter 7: Tree-based Methods and Boosting (7%)

Final Project: (30%): Students work individually or in teams to apply course concepts to a real-world dataset or problem of their choice. The project involves formulating a problem, preprocessing data, implementing machine learning methods, evaluating performance, and presenting findings. A draft proposal (5%) and a comprehensive final report (25%) are required. The report should include motivation, methodology, results, discussion, and conclusions.

Final Project Presentation: (10%): Each team delivers a presentation to the class. Presentations must cover the problem statement, data, methods, results, and key insights. Active participation in peers' presentations is required. Presentation quality and engagement will be graded.

Attendance and Participation: (10%): Regularly attending class and actively participating in discussions and activities will be rewarded with points. Each student starts with the maximum attendance and participation grade and may lose points for unexcused absences or lack of active engagement.

Bonus Reflection Assignment: (5%): At the end of the semester, students are given a bonus assignment to reflect on their learning experience in the course. Students can share their thoughts on the most valuable concepts learned, challenges faced, and how the course impacted their understanding of statistical machine learning. The reflection assignment allows students to provide feedback on the course and their own growth as learners.

Points earned	Grade
85-100+	A
75-84	B
65-74	C
50-64	D
0-49	F

Course Policies

Email: Information about the course (changes to assignments, reminders, schedules, etc.) will be distributed to students using their BlazerID email address or Canvas. Each student is required to access their UAB email account daily, as these communications represent official university business. This is a requirement for all UAB students. For UAB email account assistance, send an email to userservices@uab.edu, or call 934-3540.

Extended Absences: Attendance is fundamental to course objectives and to the integrity of this course. Courses in the Mathematics Department require a variety of activities that involve interaction with the instructor and/or interaction with other students. Excessive absences and missed assignments (more than 2 weeks) seriously jeopardize a student's ability to successfully complete the course. In the event of excessive absences, students should be prepared to officially withdraw from the course through the Registrar's Office. In cases involving medical hardships, military duty, or other serious personal situations after the withdrawal date for a course, the student may participate in the Academic Policy Appeal (accessed and submitted through Blazernet Links/Forms).

Disability support Services (DSS): UAB is committed to providing an accessible learning experience for all students. If you are a student with a disability that qualifies under Americans with Disabilities Act (ADA) and Section 504 of the Rehabilitation Act, and you require accommodations, please contact Disability Support Services for information on accommodations, registration and procedures. If you are registered with Disability Support Services, please contact DSS to discuss accommodations that may be necessary in this course. Disability Support Services can be reached at 934-4205 or www.uab.edu/dss or in the Hill Center Suite 409.

Title IX Statement: UAB is committed to providing an environment that is free from sexual misconduct, which includes gender-based assault, harassment, exploitation, dating and domestic violence, stalking, as well as discrimination based on sex, sexual orientation, gender identity, and gender expression. If you have experienced any of the aforementioned conduct we encourage you to report the incident. For more information about Title IX, policy, reporting, protections, resources and supports, please visit http://www.uab.edu/titleix for UAB's Title IX Policy, UAB's Equal Opportunity, Anti Harassment Policy and Duty to Report and Non-Retaliation Policy.

Academic Misconduct: UAB Faculty expects all members of its academic community to function according to the highest ethical and professional standards. You are expected to be aware of, and rigorously adhere to, the UAB code of conduct with regard to academic honesty and inter-personal relations.

Academic dishonesty and misconduct includes, but is not limited to, acts of abetting, cheating, plagiarism, copying homework, fabrication, and misrepresentation. Candidates are expected to honor the UAB Academic Code of Conduct as detailed in the most current UAB Student Catalog.

Add/Drop and Course Withdrawal: Drop/Add: Deadlines for adding, dropping, or withdrawing from a course and for paying tuition are published in the Academic Calendar available online. Review the Institutional Refund Policy for information on refunds for dropped courses. Withdrawal: To avoid academic penalty, a student must withdraw from a course by the withdrawal deadline shown in the academic calendar and receive a grade of W (withdrawn). Failure to attend class does not constitute a formal drop or withdrawal. Syllabus: This syllabus is subject to changes announced in class.