STATISTC 540/STATISTC 630
Introduction to Statistical Learning/Statistical Methods for Data Science
Fall 2025

Course Description:

Introduction to some modern statistical regression and classification techniques including logistic regression, nearest neighbor methods, discriminant analysis, kernel smoothing, smoothing spline, local regression, generalized additive models, decision trees, random forests, support vector machines and deep learning. Clustering methods such as K-means and hierarchical clustering will be introduced. Finally, there will also topics on resampling-based model evaluation methods and regularization-based model selection methods. The course emphasizes the mathematics behinds these methods sufficient to understand the differences among the methods as well as the practical implementation of them. *Prerequisites:* STATISTC 516 or STATISTC 608 or permission of the instructor

Administrivia:

- Time: MWF 9:05AM-9:55AM
- Location: LGRT A201
- Course Page: `maryclare.github.io/teaching.html`
- Learning Management System: Canvas
- Instructor: Maryclare Griffin
- Instructor Office: LGRT 1446, Zoom ID 848 427 5519 (as needed)
- Instructor Office Hours: T 4:00PM-5:00PM, W 10:00AM-11:00AM, Th 9:30AM-10:30AM
- Instructor E-mail: `maryclaregri@umass.edu`
- Textbooks: Introduction to Statistical Learning with Applications in `R` (`https://www.statlearning.com`) and Elements of Statistical Learning (`https://hastie.su.domains/ElemStatLearn/`)
- Computing: The computing in this course will be conducted in `R` (`https://www.r-project.org`) using RStudio (`https://rstudio.com`), both of which are freely available software available for multiple platforms.

Contacting the Instructor over Email:

I want students to contact me! However, I need your help streamlining the process. Please:

- Include 540 or 630 in the subject line to maximize the probability that I see your emails.
- Respect that the instructor responds to emails at their discretion and expect that a response may take 1 business day.
- Post questions about problem sets/homework to the appropriate Canvas forum to ensure that all students have equal access to problem sets help.

Aspirational Schedule and Key Dates:

Week 1	No Class!	9/3	9/5	Introduction, Statistical Learning, Starting Linear Regression
Week 2	9/8	9/10	9/12	More Linear Regression, Starting Classification
Week 3	9/15	9/17	9/19	Classification
Week 4	9/22	**9/24**	9/26	Exam 1 on 9/24, Resampling Methods
Week 5	9/29	10/1	10/3	Linear Model Selection/Regularization
Week 6	10/6	10/8	10/10	Beyond Linearity
Week 7	No Class!	10/15	10/17	Tree-Based Methods
Week 8	10/20	10/22	**10/24**	Exam 2 on 10/24, Guest Lectures
Week 9	10/27	10/29	10/31	Deep Learning
Week 10	11/3	11/5	11/7	Survival Analysis and Censored Data
Week 11	11/10	11/12	11/14	Unsupervised Learning
Week 12	11/17	11/19	11/21	Unsupervised Learning
Week 13	11/24	No Class!	No Class!	Unsupervised Learning
Week 14	12/1	**12/3**	12/5	Exam 3 on 12/3, Starting Presentations
Week 15	12/8	12/10	No Class!	Presentations

- 12/17: Final Project Due by 5:00PM ET.

Realized Schedule and Key Dates:

Week 1	No Class!	9/3	9/5	Introduction, Statistical Learning
Week 2	9/8	9/10	9/12	Linear Regression
Week 3	9/15	9/17	9/19	Linear Regression
Week 4	9/22	**9/24**	9/26	Exam 1 on 9/24, Classification,
Week 5	9/29	10/1	10/3	Classification
Week 6	10/6	10/8	10/10	Resampling Methods, Linear Model Selection/Regularization
Week 7	No Class!	10/15	10/17	Beyond Linearity, Tree-Based Methods
Week 8	10/20	10/22	**10/24**	Exam 2 on 10/24, Guest Lectures
Week 9	10/27	10/29	10/31	Deep Learning
Week 10	11/3	11/5	11/7	Survival Analysis and Censored Data
Week 11	11/10	11/12	11/14	Unsupervised Learning
Week 12	11/17	11/19	11/21	Unsupervised Learning
Week 13	11/24	No Class!	No Class!	Unsupervised Learning
Week 14	12/1	**12/3**	12/5	Exam 3 on 12/3, Starting Presentations
Week 15	12/8	12/10	No Class!	Presentations

- 12/17: Final Project Due by 5:00PM ET.

Grading:

- Problem Sets (Homework) 20%
- Exams 30%
- Project 30%
- Participation + Attendance 20%

Letter grades are typically as follows:

F	D*	D+*	C-*	C	C+	B-	B	B+	A-	A
<55	55+	59+	63+	67+	71+	75+	79+	83+	87+	90+

*Graduate students enrolled in undergraduate courses may receive these grades.

Problem Sets (Homework):
Problem sets/homework will be assigned weekly and due on Fridays before class via Canvas. They will be posted at least one week before they are due. The lowest problem set grade will be dropped. Late problem sets will not be accepted because we will go over the solutions immediately after submission in class. For full credit, problem sets must be submitted as an R Markdown document that has been compiled to `.pdf`. A complete submission will include two files, an `.Rmd` file and a `.pdf` file.

Exams:
There will be three equally weighted in class exams based on lecture notes and problem sets solutions.

Project:
Independent projects will be due during finals week and will include a presentation component during the final week of classes.

Participation + Attendance:
The participation grade for this course is substantial, 20%. This will be equal to $max\{i, 20\}$ where i is the number of unique lectures in which you asked or answered questions plus unique office hours you attended plus unique discussion forums you engaged with on Canvas plus the number of weeks in which you alert me to at least one typo in my posted notes.

Excused Absences:
Students who are absent due to a university-approved conflict (such as religious observance, athletic event, field trip, performance), health reasons, family illness, or other excusable extenuating circumstances remain responsible for meeting all class requirements and contacting me in a timely fashion about making up missed work. In legitimate and documented extenuating circumstances, please contact me and we will make reasonable arrangements.

University Policies:

University policies regarding Accommodations, Academic Honesty, and Title IX, apply to all courses. The policies can be found at:

`https://www.umass.edu/senate/book/non-responsible-employee-required-syllabus-statements`

Course Objectives:

At the end of the course, you should be able to:

1. Formulate real world problems in terms of statistical models and hypotheses,
2. Understand and apply data science workflow,
3. Understand mathematically the differences and advantages of various methods,
4. Implement the methods in R and write logical and coherent report.