STATISTC 540/STATISTC 630
Introduction to Statistical Learning/Statistical Methods for Data Science
Fall 2024

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 525 or STATISTC 625 or permission of the instructor

Administrivia:

- Time: TTh 10:00AM-11:15AM
- 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: TBD
- Instructor E-mail: `mgriffin@math.umass.edu` or `maryclaregri@umass.edu`
- Teaching Assistant: TBD
- Teaching Assistant Office: TBD
- Teaching Assistant Office Hours: TBD
- Teaching Assistant E-mail: `tbd@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/Teaching Assistant over Email:

We want students to contact us! However, we need your help streamlining the process. Please:

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

Schedule and Key Dates:

Week 1	9/3	9/5	Introduction, Statistical Learning, Linear Regression
Week 2	9/10	9/12	Classification
Week 3	9/17	9/19	Classification, Resampling Methods
Week 4	9/24	9/26	Linear Model Selection and Regularization
Week 5	10/1	10/3	Linear Model Selection, Beyond Linearity
Week 6	10/8	10/10	Tree-Based Methods
Week 7	No Class!	10/17	Support Vector Machines
Week 8	10/22	10/24	Deep Learning
Week 9	10/29	10/31	Deep Learning
Week 10	No Class!	11/7	Survival Analysis and Censored Data
Week 11	11/12	11/14	Unsupervised Learning
Week 12	11/19	11/21	Unsupervised Learning
Week 13	11/26	No Class!	Unsupervised Learning
Week 14	12/3	12/5	Multiple Testing
Week 15	12/10	No Class!	Poster Session

- 12/15: Final Project Due by 5:00PM ET.

Grading:

- Problem Sets (Homework) 50%
- 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 Thursdays before class via Canvas. They will be posted at least one week before they are due and there will be 12 in total. The lowest homework grade will be dropped. Late homework will not be accepted because we will go over the solutions to homework immediately after submission in class. For full credit, homework 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.

Project:
Independent projects will be due during finals week. Further details, including the structure and guidelines, will be provided as the semester progresses.

Participation + Attendance:
The participation grade for this course is substantial, 20%. This will be broken down into two components:

- $max\{i, 10\}$ where i is the number of unique lectures in which you asked questions plus office hours you attended plus discussion forum posts you engaged with on Canvas.
- One short (at most two minute, single slide) start of class presentation describing a use of a method that we have discussed or will discuss in a scientific article. Presentation dates will be assigned. Details to be provided.

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.

Academic Honesty:
Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (`http://www.umass.edu/dean_students/codeofconduct/acadhonesty/`).

Accommodation:
The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (`https://www.umass.edu/disability/`)

Title IX:
In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: `https://www.umass.edu/titleix/resources`. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

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.