M339G/M389G Syllabus

M339G/M389G: Predictive Analytics - Spring 2024 - Syllabus

COURSE-SPECIFIC INFORMATION

Welcome to M339G/M389G! Here is some information and some ground rules. Read carefully and let me know if there is anything unclear **by the twelfth day of classes, i.e., January 31st, 2024**.

Basic info

Course number. M339G/M389G (unique: 53699/54064)

Course meets. MWF 10am - 10:50am in ECJ 1.308

Instructor. Milica Čudina; my office is PMA 13.142 (2515 Speedway, Austin, TX 78712).

Email. It's best to use Canvas to email the instructor. The instructor's email address is mcudina@math.utexas.edu.

Office Hours. MWF 11am-11:50am in PMA 13.142.

Course info

Course description. This course provides an introduction to predictive modelling starting with least squares as a foundation, proceeding with classification and prediction. The emphasis will be on fitting suitable models in supervised learning, with a focus on regression and classification methods. The course includes resampling methods, simple linear regression, multiple linear regression, cross-validation, splines and tree-based methods. Some unsupervised learning methods are discussed: principal components analysis and clustering (k-means and hierarchical).

Learning outcomes.

- Expressing the idea of resampling.
- Assessing the strength of resampling procedures.
- Generalizing the simple linear regression to multiple linear regression.
- Contrasting regression and classification problems.
- Contrasting supervised and unsupervised learning.
- Assessing model accuracy in specific settings.
- Transferring ideas of linear algebra to random variable manipulations.
- Differentiating notions of accuracy.

Prerequisites. The formal prerequisite is the grade C- or better in M378K **and** M341 (or M340L, M340L-CS).

Students are also assumed to have prior programming experience preferably with `R`.

Class format and attendance. Attendance for the purposes of grading will **not** be taken. However, regular attendance is strongly recommended. In case you need to be absent, you are responsible for covering the missed material independently. Class notes will be provided on the course website. There will be **no** synchronous online option for this course. You are strongly encouraged to stay home if you are sick or contagious, not only to stop the spread of disease but also to promote your personal wellness. I view this class as a community of learners. We cannot learn effectively when we are ill. Please, take care of yourselves and your classmates.

Here are some university resources on COVID-19 and a link to the university's Exposure Action Chart.

If **students** are isolating, too sick to attend class, or experiencing another type of absence, they should:

- contact the Student Emergency Services immediately, and
- email the instructor as soon as they feel well enough to do so.

If the **instructor** is isolating, or too sick to attend class, she will do her best to change class modality to Zoom (with an alternative instructor if the situation calls for such drastic measures and if it's possible).

The class meetings consist of interactive lectures, coding demonstrations, and problem solving. In short, the course will incorporate a lot of active learning in class. Thus, if you miss class, you miss out on these learning opportunities. Please, come to class as much as possible.

Textbook. "An Introduction to Statistical Learning (with applications in `R`)" by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (Second edition)

Required devices. You will need access to a computer to be able to work on projects and homework.

Online resources.

1. **Course website**: https://mcudina.github.io/page/M339G/M339G.html. I recommend bookmarking this course site in your default browser for easy access.
2. **Canvas** will be used in this course to keep track of grades and for communication purposes. The students are responsible for the content of these announcements. The easiest way not to miss any is to turn on (i.e., not turn off) *Announcements* in their account's *Notification* menu.
3. **Ed Discussion** will be used for informal class discussion. The system is highly catered to getting you help fast and efficiently from classmates and myself. Rather than emailing questions to the instructor, I encourage you to post your questions on Ed Discussion. It is accessible via the menu on the left-hand side in Canvas.

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. 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.

Class Recordings. 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.

Assessment and grading

Homework assignments. Homework assignments will be available on the course website or in Canvas. You will be uploading your solutions using Canvas. Please, have your solutions in order and number the pages. Having read and understood this First-Day Handout in its entirety will count as the zeroth homework assignment. To get the credit, read this entire document with understanding by the homework deadline. Not handing in this assignment does not exempt you from abiding by this First-Day Handout. **The lowest three homework scores will be dropped. The homework assignments and their due dates will be announced as the term progresses.**

Projects. There will be **three** in-term group projects and one **individual final project**. The nature and content of the projects will be described in more detail as new techniques are introduced. However, every group-project will be done as part of a self-assigned group of students and require critical thinking and drawing logical conclusions.

The projects are designed to include open-ended problems which do not necessarily have a unique final answer. For that reason, there is **no** checklist-type rubric for the projects.

The formulations and due dates for the group projects will be available on the course website.

No late projects or homework are accepted except in dire circumstances at the sole discretion of the instructor.

In-term exams. There will be two in-term exams. Both will be individual and conducted in-person in our classroom. The exam coverage will be shared on the course website ahead of the exam itself. If you miss an exam due to illness or other extenuating circumstances, the final project will take the weight of the in-term exam you missed. If you miss more than one in-term exam, you are strongly encouraged to seek assistance from the Office of the Dean of Students to explore what your options are in such a dire situation.

The Final Exam. This course does not have a final exam.

The Final Project. You will be required to complete and submit a final project on a topic chosen from the list of topics provided on the course website. Your final project will be due by midnight on the Registrar-mandated date of your final exam. This semester it is **Saturday, May 4th, 2024.**

Final grade. The final grade is composed as follows:

Assignment	Percentage of final grade
Homework	14%
Group projects	36% (12% each)
In-term exams	30% (15% each)
The final project	20%

There is *no curve* in this class and the letter grades are assigned according to the following table:

A	A-	B+	B	B-	C+	C	C-	D+	D	D-
94-100	90-94	86 - 90	82 - 86	78 - 82	74 - 78	70 - 74	65 - 70	60 - 65	55 - 60	50 - 55

GENERAL, UNIVERSITY- or STATE-MANDATED INFORMATION

Drop dates. The procedure/consequences are different, depending on whether you drop before or after the 4th day of classes (01/19), and then, before or after the *main drop (Q-drop) date* (03/26). (See https://ugs.utexas.edu/vick/academic/adddrop for details)

Students with Disabilities. The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. If you have a documented disability and you need specific support as a result of your disability, please let me know as soon as possible, but definitely within the first 3 weeks of class. For more information, contact the Office of the Dean of Students at 471-6259, 471-4641 (TTY), 1-866-329- 3986 (video phone) or go to http://ddce.utexas.edu/disability/

Counseling and mental health. Counseling and other mental-health services are available from **Counseling and Mental Health Center**, Student Services Bldg (SSB), 5th Floor. (hours: M–F 8am–5pm. phone: 512 471 3515, web: http://www.cmhc.utexas.edu)

Religious holy days. Religious holy days sometimes conflict with class and examination schedules. Sections 51.911 and 51.925 of the Texas Education Code relate to absences by students and instructors for observance of religious holy days.

Section 51.911 states that a student who misses an examination, work assignment, or other project due to the observance of a religious holy day must be given an opportunity to complete the work missed within a reasonable time after the absence, provided that they have properly notified each instructor.

It is the policy of The University of Texas at Austin that the student must notify each instructor at least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holidays that fall within the first two weeks of the semester, the notice should be given on the first day of the semester. The student may not be penalized for these excused absences but the instructor may appropriately respond if the student fails to complete satisfactorily the missed assignment or examination within a reasonable time after the excused absence.

Title IX Reporting/SB 212. Texas *Senate Bill 212* requires all employees of Texas universities, including faculty, report any information to the Title IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Your instructor in a **mandatory reporter**. By law, your instructor must be fired if she does not report. Our Student Ombuds is confidential. Additionally, if you wish to speak with someone who can provide support **without** making an official report to the university, contact a confidential advocate at the Office of the Dean of Students by emailing advocate@austin.utexas.edu.

If you would like to speak with a case manager, who can provide support, resources, or academic accommodations, in the Title IX Office, please email supportandresources@austin.utexas.edu. Case managers can also provide support, resources, and accommodations for pregnant, nursing, and parenting students.

For more information about reporting options and resources, please visit: https://titleix.utexas.edu, contact the Title IX Office via email at titleix@austin.utexas.edu, or call 512-471-0419.

Sanger Learning Center. All students are welcome to take advantage of Sanger Center's classes and workshops, private learning specialist appointments, peer academic coaching, and tutoring for more than 70 courses in 15 different subject areas. For more information, please visit http://www.utexas.edu/ugs/slc or call 512-471-3614 (JES A332).

Important Safety Information. If you have concerns about the safety or behavior of fellow students, TAs or Professors, call BCAL (the Behavior Concerns Advice Line): 512-232-5050. Your call can be anonymous.

Further information about (campus) safety and security can be obtained from the Office of Campus Safety and Security, 512-471-5767, http://www.utexas.edu/safety/

Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.

- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Link to information regarding emergency evacuation routes and emergency procedures can be found at: http://www.utexas.edu/emergency

Academic (dis)Honesty. Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on academic dishonesty will be strictly enforced. For further information, please visit the Student Conduct and Academic Integrity website at: http://deanofstudents.utexas.edu/conduct For a more detailed document, please consult: https://catalog.utexas.edu/general-information/appendices/appendix-c/student-conduct-and-academic-integrity/ Please, pay particular attention to the section on *plagiarism.*

This syllabus is subject to change. If you have to miss class, please make sure to check in with a classmate to learn of any updates that were made in your absence.

The SCHEDULE of CLASSES

Date	Weekday	Topic
1/17/2024	Wed	Orientation. Rmd.
1/19/2024	Fri	Resampling Methods (Sampling Distribution and Univariate Bootstrap).
1/22/2024	Mon	Resampling Methods (Bootstrap, cont'd).
1/24/2024	Wed	Resampling Methods (Sampling Distribution of the Slope Coefficient).
1/26/2024	Fri	Simple Regression: Cross Validation
1/29/2024	Mon	Simple Regression: Prediction Interval
1/31/2024	Wed	Multiple Regression: Cross Validation
2/2/2024	Fri	Multiple Regression (cont'd) and Comments on Data Pre-processing
2/5/2024	Mon	Multiple Regression: Splines
2/7/2024	Wed	The Trade-Off Between Prediction Accuracy and Model Interpretability
2/9/2024	Fri	Supervised vs Unsupervised Learning
2/12/2024	Mon	Regression vs Classification Problems
2/14/2024	Wed	Assessing Model Accuracy: Measuring the Quality of Fit
2/16/2024	Fri	Assessing Model Accuracy: The Bias-Variance Trade-Off
2/19/2024	Mon	Collinearity
2/21/2024	Wed	Singular Value Decomposition
2/23/2024	Fri	Principal Component Analysis (Screening multivariate data)
2/26/2024	Mon	PCA and Collinearity
2/28/2024	Wed	PCA and Clustering
3/1/2024	Fri	PCA by Lagrange Multipliers
3/4/2024	Mon	Classification: The Logistic Model and KNN
3/6/2024	Wed	*In-Term One*
3/8/2024	Fri	The Logistic Model: The Confusion Matrix
3/18/2024	Mon	Classification: Linear Discriminant Analysis
3/20/2024	Wed	Linear Discriminant Analysis
3/22/2024	Fri	ROC and AUC and Discriminant Analysis
3/25/2024	Mon	Tree-Based Methods (Classification Trees)
3/27/2024	Wed	Classification Trees (Pruning)
3/29/2024	Fri	Tree-Based Methods (Bagging)
4/1/2024	Mon	Tree-Based Methods (Bagging; cont'd)
4/3/2024	Wed	Random Forest
4/5/2024	Fri	Random Forest (cont'd)
4/8/2024	Mon	**Solar eclipse.**
4/10/2024	Wed	Boosting.
4/12/2024	Fri	Maximal Margin Classifier
4/15/2024	Mon	Support Vector Classifiers
4/17/2024	Wed	Support Vector Machines
4/19/2024	Fri	K-Means Clustering
4/22/2024	Mon	A Quick Review of Hypothesis Testing
4/24/2024	Wed	*In-Term Two*
4/26/2024	Fri	The Challenge of Multiple Testing
4/29/2024	Mon	The False Discovery Rate