STAT 656 – Applied Analytics – Fall 2020

Instructor

	Office	Email	Office Hours
Darren Homrighausen	Blocker 450a	darrenho@tamu.edu[1]	See Canvas

Lecture (600)	MWF	2:55 – 3:45 pm	Online
Q&A (Section 700)	Friday	4:30 – 5:30 pm	Online

Required Text *Applied Predictive Modeling*
Kuhn & Johnson

Web Sites Canvas

Prerequisites A class that addresses multiple linear regression and some experience programming, particularly in R, will be very helpful. We will cover the relevant ideas in class as a review, but be prepared to work very hard if this is your first exposure to the topics.

Important Please direct questions about the course materials to the discussion board. Only email me about administrative issues.

Time Zone: All times stated are referencing local College Station, TX time.

Class Learning Objectives

Introduction to modern statistical learning regression and classification methods, such as neural networks, trees, and ensemble methods, from a predictive point of view. This course seeks to be very applied, discussing many details of implementing and using these methods in practice.

This class is oriented towards students that want to get an introduction to modern data analysis methods, but without a lot of the details that can make the topic seem impenetrable.

Learning Outcomes: At the conclusion of this course, successful students will be able to use R to:

- Explore, visualize and summarize data
- Build predictive models, verify assumptions, document performance, and interpret results

[1] Please only email me about course logistical issues not course content questions. Put the course # at the beginning of the subject line. I may miss your email if it doesn't include this meta data.

Administrative Remarks

Discussion Board

Discussion Board. Post any course content related questions to the discussion board. I will be monitoring it regularly, though we may let discussion threads linger a bit before intervening. The discussion board can be an invaluable tool for getting the most out of the class. However, some comments on discussion board etiquette are in order:

- Be considerate to others (respectful language, no sarcasm).
- Before posting a question, check that it (or a related question) has not already been posted. If it has, then use the existing thread for further questions or discussion. (This is important for maintaining the integrity of the board)
- Ask specific questions along with a reproducible example if it's a code question. General statements/questions aren't helpful nor likely to get helpful responses.

Software

R. In this class you will be provided the opportunity to work with the `R` statistical package. It is a free and widely used statistical computing platform. It is moderately well documented and has a vast user community (download it from `www.r-project.org`). This is the primary software for this class. Though `R` does have memory and speed constraints when confronted with large problems, we will see that it is actually still quite effective.

A convenient interface for exploring `R` is `Rstudio`, which is both an interpreter and an IDE. Some `R` resources:

- An Introduction to R
- Intro to R (video series by Google Developers)
- R studio (by RStudio)
- Simpliest introduction to R

Homework, Quizzes, Exams, and Grading Scheme

Homework. Submit all homeworks online via Canvas. As the homework assignments will involve writing a combination of code and written prose, it is mandatory that they be completed in the R Markdown format and knitted into an .html or .pdf for submission. **Submit both the .Rmd and knitted files. Homeworks in any other format will not be accepted.**

The homework problems will generally be graded using a coarse (usually a 5 point) grading scale. Occasionally, we will select a homework problem for more thorough grading. Solutions will be posted for comparing your work.

Penalties will be assessed if you

- include excess output. Excess output comes from printing out objects or running functions that you don't need in order to answer the question(s) at hand.
- turn in a submission that is essentially identical to someone else's. Collaboration is fine, but don't submit identical (or even nearly so) work.

Homework assignments that are turned in late will be assessed a 30% penalty, regardless of the reason they are late! No homeworks will be accepted after 24 hours following the original submission deadline. Feel free to discuss homework assignments with others, but realize that the work you hand in must be your own.

Homeworks will be equally weighted and comprise **50%** of the course grade (the lowest homework score will be dropped to account for any submission issues). There will be 6 homeworks, due roughly every other week at 11:59 pm on Sunday night.

Quizzes. There will be regular quizzes issued via Canvas. These quizzes should be completed alone, but are otherwise open book/notes. Quizzes will comprise **10%** of the course grade. I will drop the lowest quiz grade to compensate for any issue that might arise. **Note:** No make-ups nor extensions will be allowed.

Exams. There will be no midterm/final. There will be a project instead.

Project. There will be course-long project worth **40%** of the course grade. Project options will be listed on the course website. More details will be posted on the website for each of the following tasks:

- **Sept 6th by 11:59 pm:** Each person must submit a 1 minute video recording of you describing the problem and what you propose to do. Also, form groups of 2 to 3 for the rest of the project tasks of the semester
- **Oct. 11th by 11:59 pm:** Each group must submit a 5 minute video explaining the project and how you propose to solve it. Include what you have learned in what you have done so far.
- **Nov. 15th by 11:59 pm:** Each group must submit a 10 minute video in detail describing a method/technique that you are using for your project. Explain this as if I was a manager familiar with basic math/stats but not an expert.
- **Dec. 4th by 1:30 pm:** Each group must submit a 15 minute video, with each group member presenting at least 5 minutes, describing the entirety of the project. As well, submit a less than 2 page (or 1000 word) knitted R markdown file describing the key points.

Grading Scheme. This course will follow the usual grading scheme of: A=90-100%, B=80-89%, etc.

Section Number. It is extremely important that you know your section number. Section 600 students are considered on campus and Section 700 students are considered off campus (regardless of your physical location).

Covid-19 Announcement

Campus Safety Measures. To promote public safety and protect students, faculty, and staff during the coronavirus pandemic, Texas A&M University has adopted policies and practices for the Fall 2020 academic term to limit virus transmission. Students must observe the following practices while participating in face-to-face courses and course-related activities (office hours, help sessions, transitioning to and between classes, study spaces, academic services, etc.):

- Self-monitoring—Students should follow CDC recommendations for self-monitoring. Students who have a fever or exhibit symptoms of COVID-19 should participate in class remotely and should not participate in face-to-face instruction.
- Face Coverings—Face coverings (cloth face covering, surgical mask, etc.) must be properly worn in all non-private spaces including classrooms, teaching laboratories, common spaces such as lobbies and hallways, public study spaces, libraries, academic resource and support offices, and outdoor spaces where 6 feet of physical distancing is difficult to reliably maintain. Description of face coverings and additional guidance are provided in the Face Covering policy and Frequently Asked Questions (FAQ) available on the Provost website.
- Physical Distancing—Physical distancing must be maintained between students, instructors, and others in course and course-related activities.
- Classroom Ingress/Egress—Students must follow marked pathways for entering and exiting classrooms and other teaching spaces. Leave classrooms promptly after course activities have concluded. Do not congregate in hallways and maintain 6-foot physical distancing when waiting to enter classrooms and other instructional spaces.
- To attend a face-to-face class, students must wear a face covering (or a face shield if they have an exemption letter). If a student refuses to wear a face covering, the instructor should ask the student to leave and join the class remotely. If the student does not leave the class, the faculty member should report that student to the Student Conduct office for sanctions. Additionally, the faculty member may choose to teach that day's class remotely for all students.

Personal Illness and Quarantine. Students required to quarantine must participate in courses and course-related activities remotely and must not attend face-to-face course activities. Students should notify their instructors of the quarantine requirement. Students under quarantine are expected to participate in courses and complete graded work unless they have symptoms that are too severe to participate in course activities.

Students experiencing personal injury or Illness that is too severe for the student to attend class qualify for an excused absence (See Student Rule 7, Section 7.2.2.) To receive an excused absence, students must comply with the documentation and notification guidelines outlined in Student Rule 7. While Student Rule 7, Section 7.3.2.1, indicates a medical confirmation note from the student's medical provider is preferred, for Fall 2020 only, students may use the Explanatory Statement for Absence from Class form in lieu of a medical confirmation. Students must submit the Explanatory Statement for Absence from Class within two business days after the last date of absence.

Class Schedule

Order	Chapter(s)	Topics
0	Chapters 1.3, 1.6, Appendix B.1 - B.8	Course overview including terminology and notation. General coding practice, R, split-apply-combine
1	Chapters 3 - 5 & 11	Data preprocessing, cross validation, & bootstrap. Parallelism . The Caret package in R. Measuring performance in regression & classification.
2	Chapters 6.1 - 6.2 & 12.1 - 12.3	Multiple linear regression, Logistic regression, Linear discriminant Analysis
3	Chapters 6.4 & 12.5	Penalized/regularized models
4	Chapters 6.4-6.5 & 12.7	Additional topics on penalized/regularized models such as relaxed lasso and sparse data
5	Chapters 7.1-7.5 & 13.1-13.2 & 13.4-13.7	Nonlinear methods including Neural Networks, Support vector machines, Splines, K-Nearest neighbors
6	Chapters 8.1-8.3 & 14.1-14.2	Decision trees
7	Chapters 8.4-8.5 & 14.3-14.4	Bagging and Random Forest
8	Chapters 8.6 & 14.5	Boosting
9	Notes	A very brief introduction to text processing
10	Chapters 16	Classification with severe class imbalance

Note: The reason the chapters are so jumbled is that the book treats regression and classification separately. I feel for this class and the level at which it is being taught, that treating regression and classification simultaneously is more efficient and streamlined. This schedule is a general guide.

Important Topics

Attendance Policy. The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments. Please refer to Student Rule 7 in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Make Up Work Policy. Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student's grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor. Please refer to Student Rule 7 in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor" (Student Rule 7, Section 7.4.1).

"The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence" (Student Rule 7, Section 7.4.2).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code. (See Student Rule 24.)

Disability Resources. Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact Disability Resources in the Student Services Building or at (979) 845-1637 or visit `disability.tamu.edu`. Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

Honor Code. "An Aggie does not lie, cheat or steal, or tolerate those who do."

"Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one's work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case" (Section 20.1.2.3, Student Rule 20).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at `http://aggiehonor.tamu.edu`

Title IX and Statement on Limits to Confidentiality. Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit gender-based discrimination and sexual harassment, including sexual assault, sexual exploitation, domestic violence, dating violence, and stalking.

With the exception of some medical and mental health providers, all university employees (including full and part-time faculty, staff, paid graduate assistants, student workers, etc.) are Mandatory Reporters and must report to the Title IX Office if the employee experiences, observes, or becomes aware of an incident that meets the following conditions (see University Rule 08.01.01.M1):

- The incident is reasonably believed to be discrimination or harassment.
- The incident is alleged to have been committed by or against a person who, at the time of the incident, was (1) a student enrolled at the University or (2) an employee of the University.

Mandatory Reporters must file a report regardless of how the information comes to their attention – including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Although Mandatory Reporters must file a report, in most instances, you will be able to control how the

report is handled, including whether or not to pursue a formal investigation. The University's goal is to make sure you are aware of the range of options available to you and to ensure access to the resources you need.

Students wishing to discuss concerns in a confidential setting are encouraged to make an appointment with Counseling and Psychological Services (CAPS).

Students can learn more about filing a report, accessing supportive resources, and navigating the Title IX investigation and resolution process on the University's Title IX webpage.

Statement on Mental Health and Wellness. Texas A&M University recognizes that mental health and wellness are critical factors that influence a student's academic success and overall wellbeing. Students are encouraged to engage in proper self-care by utilizing the resources and services available from Counseling & Psychological Services (CAPS). Students who need someone to talk to can call the TAMU Helpline (979-845-2700) from 4:00 p.m. to 8:00 a.m. weekdays and 24 hours on weekends. 24-hour emergency help is also available through the National Suicide Prevention Hotline (800-273-8255) or at suicidepreventionlifeline.org.