

STAT 421: Section 500: MACHINE LEARNING

Department of Statistics, Texas A&M, SPRING 2022

Version: March 14, 2022.

The syllabus is subject to change, so look for the most recent version.

Instructional Assistant Professor

Name: Dr. Shahina Rahman

Email: srahman@stat.tamu.edu

NO EMAIL STRICTLY BETWEEN 6PM - 6AM

I do not reply to emails within this time frame.

Teaching Assistant

Huiya Zhou

huiya@stat.tamu.edu

Lectures

Tuesday and Thursday, 11:10 am -12:25 pm

BIOC 457

In person

Recitations/Labs

Friday, 3:00 - 4:00pm

TA: Zoom Recitation link

(log in with TAMU id)

Office Hours

Tuesday, 2:00 - 4:00pm

BLOC 437

In person

Virtual Office Hours

Friday, 1:00 - 3:00pm

TA: Zoom Office hours link

(log in with TAMU id)

Course Description

- The Tentative coverage of topics starting from **Linear Models, Logistic Regression, LDA, QDA, Neural Networks, Cross Validation, SVM, Tree based Regression and Classification, PCA, Clustering** is given at the end of the syllabus.
- The goal of this class is to provide a grounding in the basic concepts of pattern recognition and machine learning. It is aimed at **advanced undergraduates and masters or first/second year PhD students**, as well as researchers and practitioners. In this course you will be trained using models to analyze data — as data summaries to tell important stories, as predictive instruments for future use, and as tools for scientific inference.
- **The student is expected to know basic Python which is already taught in the previous introductory course STAT 335 : An Introduction to Datascience.** We will NOT teach you basic Python.
- **We only expect that you know how to use basic Python (numpy, looping, defining functions and lambdas and pandas to handle real life data).**
- We also expect the students will know how to use **differentiation** and **integration** of functions and basic Matrix Algebra.

- Students are expected to know basic Probability.
- Many new problems in science, industry, arts and entertainment require traditional and non-traditional forms of data analysis. In this course, you will learn how to use a set of both supervised and unsupervised learning methods for modern data mining.
- Most importantly we will focus on **why** to use many important methods or tools. The **assumptions** under which these tools works. In what way these tools are **different**. And **how to compute and implement** them in real data, and when *not* to use it. Most importantly, you will learn how to think about and model data analysis problems.
- The Calendar information is tentative and subject to change. Any changes will be notified in the Canvas and by email. And you are always free to ask before making any elaborate plans for vacation or wrapping up. **The Exam dates are FIXED.** So you should prioritize them !

Prerequisites

STAT 211 or equivalent or ECEN 303, probability, multivariate calculus, linear algebra, and statistical learning.

Grading Policies

Your grade will be determined as follows:

- **Homework and Recitation: 50%.** This will be due on ecampus on every Friday, at 11:59pm. **Recitation** will be graded 100% only upon completion. Recitation need not be 100% correct.
- **Midterm: 20%** March, 1, Tuesday, 2022, 11:10 am - 12:25 pm.
- **Final Exam: 30%** May 5, Thursday, 2022, 3:00 - 5:00 pm (As per the University Calendar).

Grading Scale

A	90-100
B	80-89
C	70-79
D	60-69
F	<60

Important Resources

- **CANVAS :** All lecture notes and homeworks will be done through CANVAS. <https://canvas.tamu.edu>. HW and Exam Grades will also be posted on CANVAS.
- **CANVAS Announcement :** You need to check the CANVAS Announcement and your tamu emails for updates on important announcement. MAKE SURE you **TURN ON** your **NOTIFICATION** for Announcements. It is important that you keep yourself informed with latest changes.
- **Github :** My Github Link. Git hub is a social media for coders. You must register to open an account on Github and start following my account. Important datasets and project information will be given here.

- **TAMU Google drive:** All lectures will be recorded and will be uploaded on the TAMU google drive which you can only access through your TAMU email id. Here is our course **Google Drive Link**
- **Python 3 Software:** You can download this software for free at the following website: <https://docs.anaconda.com/anaconda/install/>. Once you install the “Anaconda Navigator” app, you have to launch the “Jupyter notebook”. TA will be helping in the first LAB on first Friday to install these necessary module for the courses. She will also upload a video on the Course Google drive as well.
- **Introduction to Python in Datacamp** The link → Datacamp Link. Start taking the course **Intro to R** immediately.
- **Textbooks MUST** *An Introduction to Statistical Learning*. James, G., Witten, D. Hastie, T. and Tibshirani, R, Springer, 2014.
The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition. Springer, 2009. Hastie, T., Tibshirani, R. and Friedman, J.
Python for Probability, Statistics, and Machine Learning, 2nd Edition. Springer, 2019 Unpingco, J.
- **Python Modules** Standard modules such as Numpy and Matplotlib, as well as the datascience module must be installed. TA will be helping you.
- **Supporting Student Mental Health** With so much currently going on in the world and in their lives, students’ mental health can play a major role in their academic motivation, engagement, and achievement. Some of the books really helped me focus in this chaotic world is by Cal Newport, **Deep Work** and **Getting Things done**, by David Allen. And also I will try to input some strategies during my lecture to increase your focus towards your well-being.

Homework Format

- All Homeworks and Recitation must be submitted in Jupyter notebook and must be uploaded to the CANVAS.
- For Computational or code party they must use the Code cell.
- For theoretical answer, they must use the text part of Jupyter notebook.
- For writing formulas or equation, students are encouraged to add the screenshots of their solution. Hand-written documents must be CLEAR for TA to understand.
- Students must thoroughly explain their answers how they arrived at their solutions.
- Homeworks not adhering to these requirements will receive no credit.

Attendance: This class is demanding on homeworks and recitation. It’s students responsibility’s to participate and learn various machine learning tools. The classes will be held in person so there will be **NO video recording** that will be uploaded on CANVAS. In case you miss the class, it’s your responsibility to do the self learning. If you are sick, I will be happy to discuss with you the materials in person in my office hours or by appointment. On the

Lecture slides there will be some empty spaces which will be filled during the lectures after discussion participation from students. Make sure you know this before you decide not to show up in the class.

Other Remarks:

- It is difficult to discuss mathematics/statistics questions by email. You are encouraged to ask your technical questions during/after class.
- Write STAT421 in the subject heading of all email correspondence. I teach another course and I get email from students from other course. It will be hard to relate to your concerns if you do not identify your course number to me.
- Students are expected to make a serious attempt to solve an exercise or understanding a method before asking the Prof/TA.
- Group studying is encouraged but not COPYING Each student must submit his/her original work.
- Students copying homework will be heavily penalized. TA can deduct flat 50% of the homework score of either students who involved in the copying homeworks.
- Students copying Exams 50% flat reduction on two similar Exam works.

Course Policies

Exam Policy: For Online Exam, All Exams will be downloaded from eCampus.
For Online Exam, Exam solutions *must* be scanned into a single portable document format (PDF) file only and uploaded to eCampus.
You should be identified on the initial page with your PRINTED Name, Course and UIN.
Your exam solutions must be your own work, consistent with the university on academic integrity.
Each exam will be comprehensive, cumulative and open book. You will be allowed to use a self generated formula sheet. As a part of your solutions to problems, you will need to:

- Show all your work. This does not necessarily mean showing every individual algebraic or calculus step – but it must be clear what those steps are.
- Clearly identify the solution to all problems.

Copies of old exams will be available for you to review under **Review Materials for Exams** folder on eCampus.

Makeup Policy:

- If you missed a homework assignment or exam, see the university rule on Attendance website Rule 7: <http://student-rules.tamu.edu/rule07>
- If you fail to submit a homework assignment by the due date because of a university excused absence or due to illness or circumstances beyond your control, notify me in writing or by email (before, if feasible, otherwise within two working days after you return). If your absence is approved, I will notify you on how you may make up the missed assignment.
- If you must miss an exam because of a university excused absence or due to illness or circumstances beyond your control, notify me in writing or by email (before, if feasible, otherwise within two working days after you return). If your absence is approved, I will notify you on how you may make up the exam.
- If you miss a homework assignment or an exam and your reason for missing the assignment or exam is not accepted, then you will receive a score of 0 for the assignment or exam.
- A temporary grade of I (Incomplete) at the end of a semester indicates that the student has **completed the course with the exception of a major quiz, final exam, or other work**. The instructor shall give this grade only when the deficiency is due to an authorized absence or other cause beyond the control of the student.

Course Policies (continued)

Disabilities Help:

Americans with Disabilities Act (ADA) Policy Statement - The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services, currently located in the Disability Services building at the Student Services at White Creek complex on west campus or call 979-845-1637. For additional information visit <http://disability.tamu.edu/>

Academic Integrity:

“An Aggie does not lie, cheat, or steal or tolerate those who do.” You are expected to maintain the highest integrity in your work for this class, consistent with the university rules on academic integrity. This includes not passing off anyone else’s work as your own, even with their permission. Please see the homework and exam policies below for specifics. Further information at <http://aggiehonor.tamu.edu>

Copyright:

Faculty members own copyright in their educational work at Texas A&M University, as stated in the Texas A&M University System Policy for Intellectual Property Management and Commercialization (<http://policies.tamus.edu/17-01.pdf>).

Students are not allowed to post or share any materials created by a faculty member unless given permission by that faculty member. This includes but is not limited to homework assignments, homework solutions, exams, exam solutions, lecture notes and any other supplemental materials.

Any violation of this copyright policy could result in disciplinary actions as described in

Student Rule 20.2: Procedures in Scholastic Dishonesty Cases and

Student Rule 20.1.2.3.1 Cheating:

Intentionally using or attempting to use unauthorized materials, information, notes, study aids or other devices or materials in any academic exercise. Unauthorized materials may include anything or anyone that gives a student assistance and has not been specifically approved in advance by the instructor.

Texas A&M complies fully with the Digital Millennium Copyright Act (“DMCA”). Users of the Texas A&M network found to have engaged in repeated infringement of copyright are subject to termination of their network access and may be reported to the appropriate Dean or Human Resources officer for disciplinary action.

Please see TAMU’s Copyright Infringement Policies and Sanctions Notification for additional information.

[https://security.tamu.edu/protect_myself/
Copyright_Infringement_Policies_and_Sanctions_Notifications.php](https://security.tamu.edu/protect_myself/Copyright_Infringement_Policies_and_Sanctions_Notifications.php)

COVID-19 Temporary Amendment to Minimum Syllabus Requirements

Campus Safety Measures:

The Faculty Senate temporarily added the following statements to the minimum syllabus requirements in Fall 2020 as part of the university's COVID-19 response.

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.

Course Policies (continued)

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.

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.

Course Policy: Statement of Respect:

- *Preferred Name and Preferred Gender Pronouns*

“Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student’s legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records.”

- *Policies and Expectations*

This course may involve participation in discussions and activities. We each have different experiences and that influence our perspectives of the world. You may feel uncomfortable or disagree with certain ideas or opinions expressed by others or with certain topics in the class. You may also find you share perspectives or experiences with others. The classroom should be a lively, interactive and comfortable place where information is shared ideas tested and issues debated. Comments shared within this course are to remain confidential to the audience of this course; do not copy or share statements or writings from this course with others not in this course section. As an undergraduate student you should expect the professor of the course to be prepared for each course, to treat each student with respect and compassion. Students will in turn come to each class prepared, complete assignments on time and will adhere to the university’s academic standards including those governing academic dishonesty including cheating plagiarism or fraud. The following link will provide further guidance on the Universities student rules and policies: <http://student-rules.tamu.edu/>

- *Non-Discrimination Policy*

Texas A&M is committed to the fundamental principles of academic freedom, equality of opportunity and human dignity. To fulfill its multiple missions as an institution of higher learning, Texas A&M encourages a climate that values and nurtures collegiality, diversity, pluralism and the uniqueness of the individual within our state, nation and world. All decisions and actions involving students and employees should be based on applicable law and individual merit.

Texas A&M University, in accordance with applicable federal and state law, prohibits discrimination, including harassment, on the basis of race, color, national or ethnic origin, religion, sex, disability, age, sexual orientation, or veteran status.

Course Policy: Statement of Respect:

- It is the policy of Texas A&M University not to discriminate based on gender, age, disability, race, color, religion, marital status, veteran's status, national or ethnic origin, or sexual orientation. Harassment of a student in class, i.e., a pattern of behavior directed against a particular student with the intent of humiliating or intimidating that student will not be tolerated. The mere expression of one's ideas is not harassment and is fully protected by academic freedom, but personal harassment of individual students is not permitted.

- *Professionalism and Respect*

You are expected to treat your instructor and all other participants in the course with courtesy and respect. Your comments to others should be factual, constructive, and free from harassing statements. You are encouraged to disagree with other students, but such disagreements need to be based upon facts and documentation (rather than prejudices and personalities).

Unprofessional or disrespectful conduct will result in a lower grade on an assignment. Warnings will not be given; part of the learning process in this course is respectful engagement of ideas with others.

Students will need to contribute in intelligent, positive, and constructive manners within the course. Behaviors that are abusive, disruptive, or harassing may result in disciplinary actions as specified within the Student Rules. The student rules can be found at the website page: <http://student-rules.tamu.edu/statement>

Sensitive or controversial topics might be discussed in this course. Comments posted within this course are to remain confidential to the audience of this course; do not copy or share messages or writings from this course with others not in this course section.

Students in this course are also responsible for being familiar with the Universities student rules and policies: <http://student-rules.tamu.edu/>

Table 1: Tentative Spring 2022 Calendar (Subject to Change)

Week #	Class	Subject	Reading (James & Hastie)	Evaluation
Supervised				
01	Jan 18 - 20	Supervised Learning - Regression and Prediction	J 2.1-2.2	Python Tutorial in Lab, HW1 assigned
02	Jan 25 - 27	Linear Regression and Checking the Assumptions	H 2	HW1 due, HW2 assigned
03	Feb 01-03	Classification: Bayes Classifier, KNN classifier and Logistic Regression	J 4.4-4.5, H 4.3, H 6.6.3	HW2 due, HW3 assigned
04	Feb 08-10	LDA & QDA & Naive Bayes Classifier and Resampling Methods		HW3 due, HW4 assigned
05	Feb 15-17	Cross Validation and Bootstrap	J 5.1, H 7.10, J 5.2-5.3, H 7.11	HW4 due, HW5 assigned
06	Feb 22 - 24	Bootstrap and Midterm Review	H 18.1-18.3	
	Mar 1	MIDTERM EXAM		
07	Mar 3	Maximal Margin and Support Vector Classifiers	J 7.5-7.7, H 5.4-5.5, 6.1-6.3	HW5 due, HW6 assigned
08	Mar 8 - 10	Support Vector Machines	J 7.5-7.7, H 5.4-5.5, 6.1-6.3	
09	Mar 14 - 18	SPRING BREAK		
10	Mar 22-24	Trees, Bagging, XGBoosting and Random Forests	J 8.1, H 9.2, J 8.2-8.3, H 8.7, H 15.1-15.3	HW6 due, HW7 assigned
10	Mar 29-31	Neural Networks	H 9.2	
11	Apr 5-7	Deep Neural Networks		HW7 due, HW8 assigned
Unsupervised				
11	Apr 12-14	Dimension Reduction : PCA, Kernel PCA and t-SNE	J 6.3.1, 10.2 H 14.5.1-14.5.3	HW8 due, HW9 assigned
11	Apr 19-21	Clustering - Kmeans and GMM		HW9 due, HW10 assigned
12	Apr 26-28	Clustering - Hierarchical and Spectral		
13	May 3	FINAL REVIEW DAY 11	H 14.3	HW10 due, HW11 assigned on Tuesday April 26
	May 5	FINAL EXAM	Time: 3:00 - 5:00 pm	HW11 due on May 3