Machine Learning in Python
CS 5324 001C
Computer Science Department
Fall 2023

Instructor Information

Instructor: Eric Larson
Email: eclarson@smu.edu
Faculty Website: eclarson.com
Office Location: 451 Caruth
Office Hours:
Mondays 3:30-5PM, In person: Caruth 451 and Zoom: https://smu.zoom.us/j/2899688143 Email instructor if planning to attend via Zoom
Preferred Method of Contact:
Email and Office Hours
Additional Information:
Eric Larson is an Associate Professor in the department of Computer Science in the Bobby B. Lyle School of Engineering, Southern Methodist University. His main research interests are in machine learning, sensing, and signal & image processing for ubiquitous computing applications, in particular, for healthcare and security applications.

Course Details

Meeting Days/Times/Locations:

Start Date	End Date	Location	Meeting Day	Start Time	End Time
08/21/2023	12/04/2023	CARU0379	MW	11:00 AM	12:20 PM

Credit Hours: 3.00
Course Description: Introduces the processes of learning from data. Provides an overview of a number of machine learning techniques, including pre-processing, visualization, classification, and regression, used in analytics. Covers classic and contemporary learning techniques, with emphasis on artificial neural networks and deep learning methods. Material covered will be

reinforced through hands-on experience using state-of-the art tools. Class examples and assignments will come from the programming language Python. Knowledge of linear algebra, calculus, introductory algorithm analysis, statistics/probability, and an introduction to python programming is suggested. Prerequisites: C- or better in CS 2341 and MATH 3304.

Student Learning Outcomes

This course is constructed to help students design and use machine learning techniques. The course places special emphasis on learning Artificial Neural Networks. Students will hone their abilities to analyze data, visualize and explain data, and predict outcomes using various learning algorithms. While this course is designed for breadth of topics, some topics will be discussed in depth (such as neural networks and deep learning techniques). Finally, students will learn to communicate ideas about these technical areas effectively.

Topics covered include:

- Data analysis in python (scikit-learn, numpy, pandas, and jupyter)
- Visualization using matplotlib, seaborn, and interactive plots (perhaps Plotly, Altair, Bokeh, or HoloViz)
- Feature dimension reduction and feature manipulation (including PCA and UMAP)
- Introduction to common features for text and image processing
- Gradient-based optimization techniques
- Classification
    - logistic regression
    - artificial neural networks
    - wide and deep networks
    - convolutional neural networks
    - sequential neural networks (introduction only to recurrent networks and transformers)

A Note to Those Thinking of Taking the Course

The course is heavy in python programming, linear algebra, and multivariate optimization. This course is entirely project based and you are expected to work in teams. Projects will require considerable data analysis and time. There is a high expectation for this course—you are expected to stay current with the material and investigate papers/blogs/books outside of class to boost your theoretical understanding of the material. Some lectures will follow a flipped lecture

format, where you will watch videos outside of class and work on a programming assignment during traditional lecture time.

Pre-requisite knowledge for this class includes python programming, linear algebra and calculus. It is also helpful to have knowledge of introductory algorithm analysis and statistics/probability (or desire to learn quickly). I cannot stress enough that it is highly recommended to have an introduction to python programming before taking this course. Experience with databases is a bonus, but certainly not required.

This class introduces the processes of learning from data. This course provides an overview of a number of machine learning techniques (i.e., pre-processing, visualization, classification, and regression) used in analytics. We will cover a number of classic and contemporary learning techniques. The major emphasis in the course is on linear models and neural networks. All material covered will be reinforced through hands-on experience using state-of-the art tools. Class examples and assignments will come from the programming language Python. We will use APIs that include scikit-learn, tensorflow, and keras.

Assignments will require the use of the python programming language. Assignments are due in the format of “reports” that include code, visualizations, and summarize all analysis with text or video.

Required Texts and Materials

None
Required Reading Text: None

- Most information I present will come from an array of disparate sources. While many theoretical elements come from the recommended text, it is best used for review of lecture content and examples.

Recommended Reading Text (choose one from the following two):

- Python Machine Learning, by Sebastian Raschka (Highly recommended)
    - Third Edition, ISBN: 978-1789955750
    - Link: https://www.amazon.com/Python-Machine-Learning-scikit-learn-TensorFlow/dp/1789955750 (Links to an external site.)
- Hands-on Machine Learning with Scikit-learn and Tensorflow, by Aurélien Géron (Highly recommended)
    - First Edition, ISBN: 978-1491962299

    - Link: https://www.amazon.com/Hands-Machine-Learning-Scikit-Learn-TensorFlow/dp/1491962291 (Links to an external site.)

Other Optional Resource Texts:

- Neural Networks and Deep Learning, by Michael Nielsen
    - First Edition, open access
    - Link: http://neuralnetworksanddeeplearning.com (Links to an external site.)
- Introduction to Machine Learning, by Alex Smola and S.V.N. Vishwanathan
    - First Edition, ISBN:0 521 82583 0
    - http://alex.smola.org/drafts/thebook.pdf (Links to an external site.)

Grading Policies/Grading Scale

Overview

Team coding is highly encouraged in this course. All assignments can be completed as a team. For team portions of assignments, one grade will be issued to the entire team, not individually (that is, the team will sink or swim together for that portion of the assignment). This helps to prevent any issues cropping up about copying/cheating from individuals in the class. Please handle team conflicts within your team in an appropriate manner. Teams will be evaluated based upon their lab assignments and live-session assignments as follows:

- Post Class Canvas Quizzes (5% total)
- Seven Lab Assignments for 70% of the grade (10% each)
- Three live session assignments for 15% of the grade (5% each)
- Class participation for 10% of the grade
- Note: A fourth live session assignment may be used in the course and replaces 5% of the participation grade. This depends upon whether the live session recording can be completed and ready before the semester end.

Participation

Classroom students will be graded on participation using a photo assessment system and flipped module quizzes. Distance students can submit answers to “self-test” questions posed during the videos and upload them to canvas. Distance students are responsible for finding the questions in the videos, writing their answer, and uploading to canvas. Full credit will be given if all questions are uploaded through out the semester (preferably weekly). Turning in all questions at the end of

the semester will result in a zero for class participation. Remember to upload answers to self-test questions weekly here: Participation

Five of the ten participation points are from the posed in-class questions. These five points will be scored as follows:

- 5 out of 5: Answering 90% of all questions, with 50% or more of all questions correct, evenly paced throughout semester.
- 4 out of 5: Answering 90% of all questions, with <50% of all questions correct, evenly paced throughout semester.
- 3 out of 5: Answering 75% of all questions (perhaps not correct), evenly paced throughout semester.
- 2 out of 5: Answering 50% - 75% of questions (perhaps not correct), but evenly spaced throughout semester
- 1 out of 5: Answering 50% or more of questions, but not regularly spaced throughout semester (i.e., not answering questions regularly)
- 0 out of 5: Answering less than 50% of questions

The remaining five points for participation are for regularly attending class, participating in discussions, and staying current in attempting the quizzes.

Lab Assignments Grading Schema

In each lab assignment, specific deliverables are asked and should be completed to the best of your ability. Each deliverable will be worth a certain percentage of the grade and you will be graded in terms of the quality of your analysis. Comment code so that it is readable and, where possible, immediately understandable. The sum total of the these deliverables will be 90% of the points possible for each assignment. If you complete all the project deliverables satisfactorily you should expect a grade of 90%.

The remaining 10% of the points are reserved for exceptional work that is above and beyond in one or more elements of the analysis. The instructor will provide suggestions regarding exceptional work for each lab assignment. 7000 level students must complete the suggested work. 5000 level students have free reign to provide whatever additional analysis they prefer. Grading of the exceptional work does not follow a strict rubric. It is subjective to the instructor and at the instructor’s discretion.

Live-Session Assignment Grading Schema

These assignments are completed under time pressure. The instructor understands this, but also has an expectation that the students are prepared via watching videos. It is not the instructor's wish for these assignments to be “evaluations” of your learning, but, instead, be tools for

motivating you to understand the material. Use these assignments to ask questions to the instructor and get one-on-one feedback!

Curves

A dividing line for converting percentages to letter grades must be taken in order to be in line with an academic institution's grading schema. The dividing lines for this course are listed below. Even so, please do not measure your performance in this class by the letter grade you receive. Your potential employers might care very little that you got a B or C in this class if you can knowledgeably talk about machine learning systems. Also, I am happy to write recommendation letters for those that performed well in the class (regardless of your grade—performance is different than a letter grade). No extra credit will be given.

A	A-	B+	B	B-	C+	C	C-	D	F
93.000-100.00	90.000-92.999	87.000-89.999	83.000-86.999	80.000-82.999	77.000-79.999	73.000-76.999	70.000-72.999	60.000-69.999	Else

Assignment Group Descriptions

Lab Assignments

About every two weeks lab assignments will be submitted electronically. Lab assignments can be completed in teams of two or three. Late labs will not be accepted.

Lab assignments should be turned in as HTML (see bullet below). Do not include source data when uploading

1. Rendered Jupyter notebooks HTML (*.htm or *.html) rendering of the Jupyter notebooks (with all images either in a zipped directory with the master HTML file or embedded in the HTML). Please use "zip" to compress directories (i.e., try to avoid *.tar or *.rar because they are harder to view in the online system)

GitHub links of the notebook are not acceptable. Please upload the notebook to canvas. Lab assignments are posted at the beginning of the course. Most assignments are turned in following a day when formal lecture does not take place. Use this extra time to complete extended or time consuming analyses of the data. There is a high expectation for these assignments. Comment all code and explain reasoning in detail in the document.

Live-Session Assignments

Periodically during the course we will hold a "Flipped Lecture Assignment" where videos are posted/watched before the live session and we will use class time to complete a live coding assignment. The specifications for the assignment will be given at the start of session and the assignment will then be turned in at the end of session. Students will work in teams and turn in

the live-session assignments at the end of the session. Come prepared to work! Live session assignments will be given in the form of a Jupyter notebook with blank spaces for questions and blank spaces for code to be filled in. The live session assignment is completed in three parts, as follows:

1. An initial notebook with a few questions will be given out. This notebook should be answered before coming to class the day of the assignment.
2. The instructor will work out the initial problems with you during the first 15 minutes of class.
3. The remainder of the assignment will then be handed out and teams can work on the assignment until the end of class.

As a team, you will answer the questions to the best of your ability. Live-session assignments should be turned in as rendered Jupyter Notebooks (exactly like the preferred method for lab assignments). These assignments are meant to be formative and facilitate one-on-one interactions with the instructor. Ask questions often and solicit feedback on given answers from the instructor!

If working as a team, all team members will need to be present to receive a grade. An absence during a “live-session assignment” cannot be made up after the assignment has been given in session. However, certain arrangements (in restricted circumstances) can be made beforehand to take the assignment early. For example, known cleared absences like conference/work-related travel, travel related to university sanctioned extracurricular activities, etc. Exceptions will be made for unexpected circumstances such as death/hospitalization of a family member.

Post Class Quizzes

During the semester a number of post class quizzes will be administered via canvas. These will take the form of multiple choice questions regarding topics presented in lecture. Some quizzes are also attached to the flipped lecture videos. These quizzes are randomly chosen for some lecture topics and will be available on canvas periodically throughout the semester.

Generative AI will be integrated into this course

Assignments in this course have been purposefully designed to integrate Generative AI in support of the learning objectives. In class, we will discuss how Generative AI is used within this discipline, including how to navigate its potential uses and abuses, how and when to attribute sources, and other developing topics.

Course Policies

Attendance

Students taking this class (other than distance sections) are expected to attend the majority of lectures in-person. However, there is some felxibility for viewing recorded lectures for students that need do so for medical or personal reasons. This is regardless of whether a student has enrolled in a distance section or not. Please consult university policy for understanding when you can and cannot attend class via Zoom.

Virtual Etiquette

If you are joining the class via Zoom, please be respectful of the students around you. This is best exemplified through:

1. Muting yourself when not speaking
2. Turning on your video camera if you feel comfortable doing so. Please do turn on your camera if you can because this will help in your paying attention (try not to look at yourself, though, :)
3. Using the chat to ask questions or post links.
4. NOT showing or posting anything that is insensitive or inappropriate (just like being in actual class)
5. Paying attention and participating when asked.
6. Raising your hand might be difficult to see in Zoom, feel free to speak up and interrupt me if you have a burning question. I won't take it as disrespectful.
7. If you have a pet, consider having them wave at the camera (if you own cat, be sure they like you before doing this)

Cheating

Cheating of any kind such as plagiarism or direct copying is strictly prohibited and against the SMU honor code. This includes direct copying from a generative language model. However, collaboration is strongly encouraged. All lab assignments can be done as a group and turned in as a group.

A note on cheating: You can and should look online for other’s that have analyzed data in Jupyter notebooks. However, if you start with their code or use it, you should 100% make it clear where you got it and what you changed (some of the best performers in this class have taken this approach). However, the reference to other work should be included directly next to the code or visualization that you are using (that is, including a list of references at bottom of the document and not making it clear where you used parts of other's code/images/etc. WILL BE considered plagiarism). Don’t try to make it such that the cheating is “plausibly deniable.” That is not a defense for behaving unethically and could result in expulsion from the SMU program. When quoting directly from another source, please place that text in quotes and cite the quote immediately. Failure to place text in quotes from other sources will be considered cheating.

I have seen a lot of questionable copying of code/text without attribution in Computer Science in recent years. Please attribute where it came from! Many times this copying is the result of not allocating enough time to complete the assignment, or a fear that you cannot get a good grade if you let me know that you are struggling with basic concepts. It is better to try and fail in these circumstances than to turn in unattributed code. In the event of plagiarism or cheating, I will turn the matter over to SMU's honor code committee, which could result in expulsion from the University. That is a daunting consequence.

https://www.smu.edu/StudentAffairs/OfficeoftheDeanofStudents/StudentConduct/HonorCouncil

Distinction between Undergraduate (5324) and Graduate (7324) Offerings

This class is taught at the graduate level. It is expected that undergraduates taking this course will adapt their behavior to a graduate level course. However, to distinguish the offerings in a manner fitting to the undergraduate and graduate expectations, there is a difference in the rubric for lab assignments. For 5000 level undergraduates, the labs have an "exceptional credit" portion that students can choose what type of analysis they perform. The rubric gives suggestions for how to complete the exceptional credit, but it is not required that 5000 level students follow this suggestion. For graduate students (7000 level) the suggested exceptional work must be completed as specified. This will require 7000 level students to learn some methodologies on their own and display understanding of the process through exceptional work.

Distance Student Information

This class is designed to work for both distance and on-campus students. However, there are a few logistical differences for those taking the classes. The following differences for distance students are summarized in the follow list:

- **Participation**: Distance students participation is graded based upon answering and uploading answers to questions posed in the videos. Whenever the instructor asks a questions and surveys the class for responses, this is a question you should make a one sentence written response for (in a text file). As more questions are posed in the videos, simply answer these questions and upload them to Participation
- **Flipped Assignments**: Throughout the semester flipped assignments are offered in which students watch videos at home, then come to class to complete a jupyter notebook. For distance students, you can download and complete this assignment within one week of the due date. Simply turn in the assignment as "late" and I will grade it appropriately. If you wish to come to class (pending availability) for these flipped assignments, that should be fine.
- **Teams**: Distance students are encouraged to form teams and you can form a team with a mix of on-campus and distance individuals. You also have the option not to form a team, if

you prefer. If you want to reach out to team members, you can create a discussion thread on canvas or email the class

- **Videos**: Videos are available in the Panopto page and posted on learning module pages. These are typically posted within 24 hours of lecture.

Title IX and Disability Accommodations

Disability Accommodations

Students who need academic accommodations for a disability must first register with Disability Accommodations & Success Strategies (DASS). Students can call 214-768-1470 or visit smu.edu/DASS to begin the process. Once they are registered and approved, students then submit a DASS Accommodation Letter through the electronic portal, *DASS Link*, and then communicate directly with each of their instructors to make appropriate arrangements. Please note that accommodations are not retroactive, but rather require advance notice in order to implement.

Sexual Harassment

All forms of sexual harassment, including sexual assault, dating violence, domestic violence and stalking, are violations of SMU's Title IX Sexual Harassment Policy and may also violate Texas law. Students who wish to file a complaint or to receive more information about the grievance process may contact Samantha Thomas, SMU's Title IX Coordinator, at accessequity@smu.edu or 214-768-3601. Please note that faculty and staff are mandatory reporters. If students notify faculty or staff of sexual harassment, they must report it to the Title IX Coordinator. For more information about sexual harassment, including resources available to assist students, please visit smu.edu/sexualharassment.

Pregnant and Parenting Students

Under Title IX, students who are pregnant or parenting may request academic adjustments by contacting the Office of Student Advocacy and Support by calling 214-768-4564. Students seeking assistance must schedule an appointment with their professors as early as possible, present a letter from the Office of the Dean of Students, and make appropriate arrangements. Please note that academic adjustments are not retroactive and, when feasible, require advance notice to implement.

Academic Policies

Religious Observance

Religiously observant students wishing to be absent on holidays that require missing class should notify their professors in writing at the beginning of the semester and should discuss with them, in advance, acceptable ways of making up any work missed because of the absence. Click here for a list of holidays.

Medical-Related Absences

To ensure academic continuity and avoid any course penalties, students should follow procedures described by their instructors in order to be provided with appropriate modifications to assignments, deadlines, and exams.

Excused Absences for University Extracurricular Activities

Students participating in an officially sanctioned, scheduled university extracurricular activity should be given the opportunity to make up class assignments or other graded assignments that were missed as a result of their participation. It is the responsibility of the student to make arrangements for make-up work with the instructor prior to any missed scheduled examinations or other missed assignments. (See current Catalog under heading of "Academic Records/Excused Absences.")

Final Exams

Final course examinations shall be given in all courses where appropriate, and some form of final assessment is essential. Final exams and assessments must be administered as specified in the official examination schedule and cannot be administered or due during the last week of classes or during the Reading Period. Syllabi must state clearly the form of the final exam or assessment, and the due date and time must match the official SMU exam schedule. SMU policy states that all exceptions to the examination schedule may be made only upon written recommendation of the chair of the department sponsoring the course and with the concurrence of the dean of that school, who will allow exceptions only in accordance with guidelines from the Office of the Provost.

Academic Dishonesty

Students are expected to embrace and uphold the SMU Honor Code. Violations of the Honor Code will be acted upon in accordance with the policies and procedures outlined in the Mustang Student Handbook.

Student Support Services

Student Academic Success Programs

Students needing assistance with writing assignments for SMU courses may schedule an appointment with the Writing Center through Canvas. Students who would like support for subject-specific tutoring or success strategies should contact SASP, Loyd All Sports Center, Suite 202; 214-768-3648; smu.edu/sasp. Tutor schedules are available at smu.edu/tutorschedule.

Caring Community Connections Program

CCC is a resource for anyone in the SMU community to refer students of concern to the Office of the Dean of Students. The online referral form can be found at smu.edu/deanofstudentsccc. After a referral form is submitted, students will be contacted to discuss the concern, strategize options, and be connected to appropriate resources. Anyone who is unclear about what steps to take if they have concerns about students should contact the Office of the Dean of Students at 214-768-4564.

Mental Health Resources: Counseling Services and Teletherapy

Throughout the academic year, students may encounter different stressors or go through life experiences which impact their mental health and academic performance. Students who are in distress or have concerns about their mental health can schedule a same-day or next-day appointment to speak with a counselor by calling Counseling Services. Counselors are available at any time, day or night for students in crisis at this number: 214-768-2277 (then select option 2) They will be connected with a counselor immediately. Students seeking ongoing counseling should call the same number (214-768-2277, then select option 1) during normal business hours to schedule an initial appointment. SMU Teletherapy provides another free option for on-demand counseling and video appointments with a medical professional.

Campus Carry Law

In accordance with Texas Senate Bill 11, also known as the ‘campus carry’ law, and following consultation with entire University community, SMU chooses to remain a weapons-free campus. Specifically, SMU prohibits possession of weapons (either openly or in a concealed manner) on campus. For more information, please see: smu.edu/campuscarrylaw.

Canvas Course Schedule

Due Date	Assignment	Type	Points

Due Date	Assignment	Type	Points
9/10	Lab One: Exploring Table Data	Assignment	10
9/15	Q1. Dimension and Reduction	Quiz	1
9/18	Flipped Assignment: Linear Regression	Assignment	5
9/24	Lab Two: Exploring Image Data	Assignment	10
10/2	Q2. History of Neural Networks	Quiz	1
10/2	Flipped Assignment: Back Propagation	Assignment	5
10/8	Lab Three: Extending Logistic Regression	Assignment	10
10/18	Flipped Assignment: Evaluation and Validation	Assignment	5
10/22	Lab Four: The Multi-Layer Perceptron	Assignment	10
10/30	Q3. Wide and Deep Networks	Quiz	1
11/5	Lab Five: Wide and Deep Networks	Assignment	10
11/19	Lab Six: CNNs	Assignment	10
11/28	Q4. Convolutional Neural Networks	Quiz	1
11/30	Q5. Recurrent Networks	Quiz	1
12/7	Lab Seven: RNNs	Assignment	10
	Help and Question with course material	Discussion	0
	Introductions	Discussion	0
	Participation	Assignment	10

Due Date	Assignment	Type	Points
	Participation Question	Discussion	0
	Mid-Semester Feedback	Quiz	0
	Team Forming	Discussion	0
	Here is how to upload HTML files to canvas	Discussion	0
	Final Grade	Assignment	100
	Loading and Formatting Image Data	Discussion	0
	I can't open the videos of Python YouTube, can you?	Discussion	0

Course Schedule

See canvas for course schedule, assignments, flipped modules, and additional information.