








 edkrueger /
eco395m-ml     

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#)

 main eco395m-ml
/ README.md 

 edkrueger Syllabus (#1) 

16da555 · last week



209 lines (137 loc) · 16 KB

Preview

Code

Blame

Raw



ECO 395m: Machine Learning (aka Data Mining/Stat Learning)

Unique Numbers: 35039

Course website: <https://github.com/edkrueger/eco395m>

Class:

Class will meet Tuesday and Thursday **3:30pm - 5:00pm CT**.

Class will be held in person in **BRB 1.118**.

TA sessions will be held in person in **BRB 1.118**.

Recordings of lectures will be posted for a limited time on Canvas through Lectures Online. Recordings are not a substitute for in class attendance.

Instructor: Edward Krueger

- Email: edwardkrueger@utexas.edu
- Office Hours: Tuesdays and Thursdays 5:00pm - 5:45pm CT
- Office Hours Location: **BRB 1.118** (unless otherwise indicated)

Teaching Assistant: Shreeyesh Menon

- Email: shreeyesh.menon@utexas.edu
- TA In Person Office Hours: TBD
- TA Office Hours Location: TBD

Course Description and Requirements

Overview and Prerequisites

This is master's level course on Machine Learning. The objective of the course is for students to gain fundamental proficiency in Machine Learning (ML). This requires knowledge of the problems that ML is expected to solve, knowledge of the models that solve these problems and the programming skills required to implement these models on real world data. In particular, we'll learn how to select models, implement, tune and evaluate machine learning pipelines in Python (using primarily scikit-learn with other packages from Python's data ecosystem). Successful students will understand how the fundamental models work and best practices for using them.

This is an introductory course with respect to ML however core knowledge of mathematics (including algebra, probability and calculus), statistics and programming is expected. We'll occasionally use some matrix math (arithmetic, algebra and calculus) in lectures, but, most likely, not in the homeworks. We'll also occasionally reference some material from econometrics that is likely not covered in statistics courses taught outside of economics and related disciplines, but this is unlikely to present a major obstacle for students from other backgrounds. There are no formal prerequisites.

We'll cover the following topics:

- Object Oriented Programming (as it relates to ML and scikit-learn)
- Linear Regression (including some applied econometrics topics)
- KNN
- Decision/Regression Trees
- Naive Bayes
- Gradient Descent

- Logistic Regression
- Regularization (including LASSO and Ridge)
- Random Forests

Time permitting, we'll also cover some of the following:

- Linear Probability Model (LPM)
- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)
- AdaBoost
- Gradient Boosting
- Dimensionality Reduction
- Clustering Algorithms
- Practical applications of Neural Networks, LLMs and AI
- Natural Language processing

We won't cover:

- *Advanced Optimization*: We won't cover advanced optimization techniques, However, we'll cover gradient descent.
- *Neural Networks and "AI"*: However, we'll cover many of their building blocks and, time permitting, utilize them through external libraries or APIs for practical applications.
- *Statistical Inference*: We will primarily be interested in prediction problems rather than inference problems. However, we will occasionally discuss inference where it helps us understand and improve our models.
- *MLOps*: However, you'll learn the scikit-learn library and particularly the pipelines interface. If you know application development, you should be able to deploy your models.
- *Explanations Linear Algebra*: Some models require Linear Algebra for a thorough treatment. For example, Support Vector Machines (SVM) and Principal Component Analysis (PCA). We'll not be able to cover these thoroughly.

Finally, some student projects may explore some these areas.

Reading and References

Our primary reference for this class will be *An Introduction to Statistical Learning* which is freely available from the authors website: <https://www.statlearning.com> If you decide to use a physical copy, which I recommend, it is fine to use either the R or Python version. The primary content is language agnostic and motivates the material separately from its implementation in code. Supplemental chapters have self contained programming labs. We'll be following the content closely for the primary content, however we won't follow the labs.

A supplemental optional reference is *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* by Aurélien Géron which offers an approach that focuses on scikit-learn and explains the concepts in the context of the code.

The Scikit-learn documentation is an excellent resource consisting of three parts:

- [API Documentation](#) describing the options for each class and function of the library.
- [User Guide](#) describing how the algorithms work and the design decisions made in the implementations.
- [Examples](#) giving a number of examples of how to use many of the models.

If there is a concept that you do not understand, refer to the resources books. Although there are many resources, the quality of resources, especially the freely available ones, varies tremendously. There is a tremendous amount of bad advice, bad practices and just plain wrong information about coding, data and machine learning on the internet.

Software

We will use open-source software in this course Python and its packages. We'll also use git. You'll be expected to install Python and Python packages as needed. If needed, the TAs will assist during TA office hours.

In addition, we'll use GitHub. The free plan will be sufficient for this course.

We'll use Discord as our primary channel for communication for everything except for grades. Announcements will be made on Discord; it is your responsibility to check for them.

How to Succeed in This Course

During Lectures

Lectures will be fast-paced, and you must pay attention in class. When I cover programming content, it can be tempting to replicate every step I take and write the same code I write *while in the lecture*. I don't recommend this as you'll likely fall behind and miss out on new material. During non-programming portions of the course, I expect your computers to be closed. So it may be wise to consider not opening your computer at all during class.

There is no legitimate reason for you to be on your cell phones during lecture.

Outside of Class

If you'd like to go back and try to replicate the work I do in class, you will have a recording of the lectures. Your homework and group projects will provide you with an opportunity to do it yourself.

After getting the big picture from lectures, I recommend doing the reading and then the homework assignment.

Please attend office hours.

During Group Projects

A large part of the learning during this course will occur during group projects. You'll harden the skills you learned in class, learn new skills and learn how to work on real-world open-ended projects.

We'll set aside class time during group projects to work on them. It's essential that you attend class during these times. It's the only opportunity you will have access to me, the TAs, and your group members *at the same time*. We'll provide advice on scoping your projects, help you make good technical choices and help you troubleshoot complex issues.

Course Policies

Grading

Grades will be based on:

- Participation (10%)
- Homework Assignments (20%)

- Midterm Group Project (30%)
- Final Group Project (40%)

Final grades will be determined on the basis of the following rubric.

Please note: to ensure fairness, all numbers are absolute, and will not be rounded up or down at any stage. Thus a B- will be inclusive of all scores of 80.000 through 83.999. A = 94-100 A- = 90-93 B+ = 87-89 B = 84-86 B- = 80-83 C+ = 77-79 C = 74-76 C- = 70-73 D+ = 67-69 D = 64-66 D- = 60-63 F = 0-60.

Grades will be *not* be curved. Please do not ask me about extra credit or extra work to improve your grade. None will be given.

Participation

Your participation grade will be based on the extent to which you attend class and participate in your groups' meetings and your attendance during the midterm and final presentations. We won't formally take attendance outside of project work weeks, but excessive tardiness may impact your participation grade. We will take attendance during project work days and presentation days and absence will affect your participation grade.

Should you have exceptional circumstances where you cannot attend class during project work, you must make arrangements with your group and inform me of these arrangements *in advance*.

Homework Assignments

- There will be up to 10 homeworks assigned during the semester.
- Homework will not be assigned during project weeks but may be due during them.
- Due dates will be announced as homework is assigned. Homeworks will typically be due a week from when they are assigned. Late homework will *not* be accepted.
- Homeworks will be submitted in a *private* Github repository. You must invite the TA(s) and me as collaborators. We'll be able to find your homeworks if you follow the instructions in the assignments. Homework solutions may not be made public. Grades will be posted on Canvas. You will not need to submit homeworks on Canvas.

Group Projects and Group Project Presentations

- Students will form groups of around five students. (Sizes are subject to change based on enrollment.)
- Students may opt to form different groups for the midterm and final projects.
- Projects will be developed and submitted in public GitHub repositories. You are encouraged to share your projects.
- Projects will be graded based on *both* group and individual requirements. GitHub will allow us to see who makes which contributions.
- Midterm Projects will be presented in class. Each group will have 5 minutes to present a talk and 5 minutes of Q & A. (Presentation times are subject to change based on enrollment.)
- Final Projects will be presented during the final exam time slot for the course. Each group will have 10 minutes to present their project and 5 minutes of Q & A. (Presentation times are subject to change based on enrollment.)
- You must attend every group's presentation on presentation days, not just your own group's.
- If you know in advance that you have a conflict with one of the presentation dates (e.g., due to travel or religious observance) or require accommodations, please see me as soon as possible to work out an alternative.
- If you will be absent from a project presentation, you must notify me prior to the presentation at edwardkrueger@utexas.edu. Where advanced notification is not feasible, **the notification must be given by the end of the second day after the absence**. Non-excused absences will result in a zero for that portion of your grade.

University Disclosures and Policies

Statement on Academic Integrity and Conduct

The University of Texas Honor Code states:

The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community.

Each student in this course is expected to abide by the UT Honor Code and uphold academic integrity.

What this means for this course:

You are encouraged to study together and to discuss information and concepts covered in lecture and the recitation sections with other students. You can work together on homework assignments. However, this cooperation should never involve one student having possession of or copying directly from another student's work. Should such copying occur, both students involved will receive zeros for the assignment. For group projects, the Honor Code means that the work you represent as your contributions must be your own. *Git and GitHub will allow us to see your individual contributions to group projects.*

Sharing of Course Materials is Prohibited

No materials used in this class, including, but not limited to, lecture hand-outs, assessments (homework assignments), in-class materials, and review sheets, may be shared online or with anyone outside of the class - or in future classes - 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. I am well aware of the sites used for sharing materials, and 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.

ADA Notice

I am committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, you may request appropriate academic accommodations by contacting Services for Students with Disabilities (SSD) at <http://diversity.utexas.edu/diversity/>.

Harassment Reporting Requirements

Under Texas Senate Bill 212 (SB 212), the professor and TAs for this course are required to report for further investigation any information concerning incidents of sexual harassment, sexual assault, dating violence, and stalking committed by or against a UT student or employee. Federal law and university policy also requires reporting incidents of sex- and gender-based discrimination and sexual misconduct (collectively known as Title IX incidents). This means we cannot keep confidential information about any such incidents that you share with us. If you need to talk with someone who can maintain confidentiality, please contact University Health Services at <https://healthyhorns.utexas.edu/> or the UT Counseling and Mental Health Center at <https://cmhc.utexas.edu/>. You can also make an appointment with a confidential advocate by emailing advocate@austin.utexas.edu or calling (512) 232-2860. We strongly urge you to make use of these services for any needed support and to report any Title IX incidents to the Title IX Office.

Wellbeing Resources:

Grad school is hard. Take care of yourselves and others.

- The Counseling and Mental Health Center serves UT's diverse campus community by providing high quality, innovative and culturally informed mental health programs and services that enhance and support students' well-being, academic and life goals. To learn more about your counseling and mental health options, call CMHC at (512) 471-3515.
- Check out the Longhorn Wellness Center, and these self-care Virtual Mindfulness and Stress Reduction Activities.
- If you are experiencing a mental health crisis, call the CMHC Crisis Line 24/7 at (512) 471-2255.
- If you have concerns about the safety or behavior of fellow students, TAs or Professors, call BCCAL (the Behavior Concerns and COVID-19 Advice Line):

512-232-5050. Your call can be anonymous. If something doesn't feel right – it probably isn't. Trust your instincts and share your concerns.

Course Outline

Topic 1 Regression

Midterm Project

Project Work Week -- *Tuesday, March 11th* will be used for in class project work.
Midterm Project Presentations will be on *Thursday, March 13, 3:30pm-5:00pm CT*.

Topic 2 Classification

Final Project

Project Work Days -- *Tuesday, April 22nd, Thursday, April 24th* will be used for in-class project work.

Final Project Presentations will be on *Monday, May 5th, 8:00am-10:00am CT*.