COMP 7747/8747: Advanced Topics in Machine Learning

Basic Information

Time & Place: Online
Instructional mode: Online
Instructor: Xiaolei Huang
Office Hours: TBD
Contact: xiaolei.huang@memphis.edu

Course Description

Computers can collect vast amounts of data from a wide range of sources, while how to efficiently understand the data is still under exploration. The goal of this course is: **read, read, and read**, aiming to train you as an independent researcher. This course will cover state-of-the-art (STOA) concepts and techniques of machine learning (ML) approaches with a special focus on applying in natural language processing (NLP). The course starts with primary concepts and methods in machine learning. Topics include necessary concepts of deep learning, word embedding, language model, Transformer, reinforcement learning and transfer learning. The special focus is for NLP, which is an interdisciplinary field that have many applications in the other fields, such as computational social science, psychology, health, cognitive science, etc. We will cover applications of neural models in several major NLP applications. Finally, we will study how machine learning models learn biases from human language.

We will have two types of reflection writings per week:

- Reading reflection: Students will read one paper that is listed in the course slides per week and will choose **one paper** (no earlier than 2019) from 6 selected top conferences (ACL / EMNLP / NAACL / NeurIPS / ICML / ICLR) each week. Additional conference papers must be approved by the instructor before the reflection submission. Students will post reading reflections on the course platform. Each student will comment two other reflection posts within a week.
- Lecture video reflection: Students will write reflection from Week 2 and post the reflection in the course website.

Each week, we'll discuss a state-of-the art approach in natural language processing and machine learning. Each class will have three components: introduction, mathematical foundation and (NLP) applications. Students who do research talks that will follow a similar format. This course will focus on readings, research talks and projects. Students will give two presentations during the course: research talk and group project presentation.

- Students will sign up groups and will give **a research talk**. Each group (2 – 3 students) will discuss an advanced *ML-related* publication or topic in detail for 45 minutes depending on final enrollments. The selected publication must be collected from the same top ML

conferences. The talk will practice presentation skills. Students must post publication URL, abstract and title of the talk (not the publication), and a short biography for each speaker on our course platform a week before your talk.
    - Each student needs to watch the talk video and submit your evaluation. The TA will provide the form to each talk presenter/group. They will post the link together with the video.
- The **final project** will have three stages, initial team proposal, one-page midterm report and a final project report. The project aims to guide students to submit conference papers at the end of this course. We will release more detailed rubric.

Prerequisite

One of the following courses:

- COMP 7745/8745: Machine Learning
- COMP 7780/8780: Natural Language Processing
- COMP 4992/6992: Intro to NLP
- COMP 7118/8118: Topics in Data Mining

We will work extensively on the Python programming language. It is assumed that you know basic machine learning techniques (logistic regression). You will also need to be able to read recent research papers in NLP and machine learning.

Textbook (*FREE*)

- The best textbook is to read the latest publications.

Schedule (subject to change)

Week 1 (01-16): Introduction to Python, Probability, NLP

- No class on 01-15
- Self-introduction **due** at 01-21 (count as participation).

Week 2 (01-22): Frameworks and Deep Learning Basics

- Reading Reflection **due** at 01-25

Week 3 (01-29): Word Representations and Evaluation

- Reading Reflection **due** at 02-01
- Research Talk Signup **due** at 02-02

Week 4 (02-05): Language Models and Deep Sequence Modeling

- Reading Reflection **due** at 02-08

Week 5 (02-12): Transformers

- Reading Reflection **due** at 02-15

Week 6 (02-19): Transfer Learning1: Sequential and Cross-lingual

- Reading Reflection **due** at 02-22

Week 7 (02-26): Transfer Learning2: Domain Adaptation and Multitask Learning

- Reading Reflection **due** at 02-29
- Final Project proposal **due** at 02-29

Week 8 (03-04): No Lecture & Spring Break

- No Reading Reflection due to Spring Break.

Week 9 (03-11): Scientific Writings

- Reading Reflection **due** at 03-14

Week 10 (03-18): Deep Generative Modeling / Model Interpretability

- Reading Reflection **due** at 03-21

Week 11 (03-25): Machine Learning for Health

- Reading Reflection **due** at 03-28

Week 12 (04-01): Multimodal Machine Learning

- Final Project Midterm Report: **due** at 04-04
- Reading Reflection **due** at 04-04

Week 13 (04-08): Computational Social Science

- No Reading Reflection.

Week 14 (04-15): ML and NLP ethics

- No Reading Reflection.
- Research Talk Slides submission **due**: depending on your signup dates.

Week 15 (04-22): Final Presentation

- Last Day of Classes: 04-24
- No Reading Reflection.
- Final Project Presentation Video & Slides submission **due**: 04-25
- Project report submission **due**: 04-28

Evaluation

Grades will be computed as follows:

Percentage	
20%	Participation
20%	Reading Reflection
20%	Research Talk
40%	Final Project
100%	Total

97+	A+
[93 - 97)	A
[90 - 93)	A-
[87 - 90)	B+
[83 - 87)	B
[80 - 83)	B-
[77 - 80)	C+
[73 - 77)	C
[70 - 73)	C-
[60 - 70)	D
< 60	F

The "[" refers to "include", and the ")" means "exclude".

Late Policy. All late submissions are subject to a penalty of 20% per day for no more than two days. Late 1 day: 80%; Late 2 days: 60%; Late > 2 days: 0%.

Final Project Policy. Reports are due at 11:59 pm Central Standard Time indicated on the schedule. Students are required to submit **PDF** files of their research talk slides and project reports. Using LaTeX is required for the final project report. You are allowed to collaborate with other peers but copying and pasting from another student will be considered plagiarism. The final project will be a group research project on a topic of students' choices after consulting with the instructor. The final project will base on a short presentation and technical report (6 or more pages excluding references). The final project report will be due during the final exam period. More information will be posted on a separated page for the final project.

Plagiarism

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and will not be tolerated. All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students

are allowed and encouraged to discuss with each other and look up resources in the literature (including the internet) on their assignments, but appropriate references must be included for the materials consulted, and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Office of Student Conduct for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to https://www.memphis.edu/osa/students/academic-misconduct.php.

"Your written work may be submitted to Turnitin.com, or a similar electronic detection method, for an evaluation of the originality of your ideas and proper use and attribution of sources. As part of this process, you may be required to submit electronic as well as hard copies of your written work or be given other instructions to follow. By taking this course, you agree that all assignments may undergo this review process and that the assignment may be included as a source document in Turnitin.com's restricted access database solely for the purpose of detecting plagiarism in such documents. Any assignment not submitted according to the procedures given by the instructor may be penalized or may not be accepted at all." (Office of Legal Counsel, October 17, 2005).

Accommodations

Any student who anticipates physical or academic barriers based on the impact of a disability is encouraged to speak with me privately. Students with disabilities should also contact Disability Resources for Students (DRS) at 110 Wilder Tower (901-678-2880). DRS coordinates access and accommodations for students with disabilities (http://www.memphis.edu/drs/).

If you are sick, in particular with an illness that may be contagious, notify me by email but do not come to class. If you are struggling with anxiety, stress or other mental health related concerns, please consider visiting the Counseling Center or calling 901.504.6442 or 901.468.3633.