

COMP 4992/6992: Introduction to Natural Language Processing

Basic Information

Time & Place: Every Tuesday and Thursday, 11:20 am - 12:45 pm

Zoom Link:

<https://memphis.zoom.us/j/6235067187?pwd=UnZDY2NFaHJoN2p4Vm00S3BzUDIFdz09>

Instructional mode: CIM (Computer Based Interactive Media)

Instructor: Xiaolei Huang

Office Hours: Tuesday 10 am – 11 am, or by appointment (email)

Course Description

This course will cover fundamental concepts and techniques of statistical machine learning approaches to natural language processing. The course starts with primary concepts and methods for processing human language. Topics include necessary concepts of probability and statistics, language and classification model, syntax, parsing and semantics. Natural language processing 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 applications of NLP techniques. Topics include topic models, information extraction, question answering, dialog and fairness.

This course will focus on hands-on assignments and projects. The final project will have three stages, initial team proposal, one-page midterm report and a final project report. Students will also complete two in-class midterm and final exams (~30 minutes each) that will test high-level understanding of concepts.

Prerequisite

COMP 2150

COMP 4001 (optional)

We will work extensively on the Python programming language. **It is assumed that you know how to program in Python and use Unix-like operating systems (Linux, OS X).**

Textbook

- SLP: Speech and Language Processing, 3rd edition:
 - <https://web.stanford.edu/~jurafsky/slp3/>
- E_NLP: Jacob Eisenstein, Natural Language Processing (2018):

- <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>

Schedule (subject to change)

Week 1: Introduction to Python, Probability, NLP

- Date: 01-19 & 01-21
- Reading:
 - E_NLP, Chapter 1.
 - E_NLP, Appendix A, page 475 – 482.
- Homework 0:
 - Set up Python development environment.
 - Basic probability warmup

Week 2: Language Model and Naïve Bayes

- Date: 01-26 & 01-28
- Reading:
 - SLP, Chapter 3.
 - SLP, Chapter 4.1 – 4.6.
 - [Chen & Goodman](#)
- Homework 1: start at 01-26
- Homework 0: **due** at 01-29

Week 3: Classification: Logistic Regression

- Date: 02-02 & 02-04
- Reading:
 - SLP, Chapter 5.
- Homework 1: **due** at 02-05

Week 4: Classification: Feature Engineering and Evaluation

- Date: 02-09 & 02-11
- Reading:
 - SLP, Chapter 4.7 – 4.10.
 - SLP, Chapter 6.1 – 6.6.
 - E_NLP, Chapter 4.
- Homework 2: start at 02-09

Week 5: Hidden Markov Models / Part of Speech Tagging

- Date: 02-16 & 02-18
- Reading:

- SLP, Chapter 8.2 & 8.5
- [HMM](#)
- Homework 2: **due** at 02-19

Week 6: Context-free Parsing

- Date: 02-23 & 02-25
- Reading:
 - SLP, Chapter 12, 13 & 14
 - E_NLP, Chapter 10
- Homework 3 starts at 02-22

Week 7: Word Embedding 1

- Date: 03-02 & 03-04
- Reading:
 - SLP, Chapter 6
 - E_NLP, Chapter 14
- Homework 3 **due** at 03-05

Week 8: Midterm Exam and Deep Learning Basics

- Date: 03-11, No class on 03-09
- Reading:
 - SLP, Chapter 7
 - [PyTorch tutorials](#)
- Midterm (take-home) Exam: 03-11

Week 9: Deep Neural Networks

- Date: 03-16 & 03-18
- Reading:
 - SLP, Chapter 9
- Homework 4: start at 03-18

Week 10: Transfer Learning and Domain Adaptation

- Date: 03-23 & 03-25
- Reading:
 - E_NLP, Chapter 5.4
 - [Ruder, Chapter 3](#)
- Project proposal **due** at 03~23

Week 11: Computational Social Science

- Date: 03-30 & 04-01
- Reading:
 - [Garg et. al. 2018](#)
 - [Social Monitoring for Public Health](#)
- Homework 4: **due** at 04-02

Week 12: Information Extraction

- Date: 04-06, No class on 04-08
- Reading:
 - SLP, Chapter 17

Week 13: Question Answering and Dialogs

- Date: 04-13 & 04-15
- Reading:
 - SLP, Chapter 23
 - SLP, Chapter 24

Week 14: Machine Translation and NLP ethics

- Date: 04-20 & 04-22
- Reading:
 - MT:
 - E_NLP, Chapter 18.1 & 18.2
 - [Papineni et al. \(2002\)](#),
 - [Collins' Notes on the IBM Models](#)
 - Ethics:
 - [Hovy and Spruit \(2016\)](#)
 - [Shah et al. \(2019\)](#)

Week 15: Final Presentation

- Date: 04-27

Week 16: Exam and Project Report

- Final (take-home) Exam Date: 04-30
- Final Project Report Due: 05-04

Evaluation

Grades will be computed as follows:

Percentage	
40%	Homework
20%	Exams

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”.

It is possible to earn extra credits by going above and beyond the expectations of the assignments, exams or final project. 6000-level students are expected to have extra challenges in the homework and exams.

Late Policy. There will be in total four no-penalty late days for students to freely distribute to the four homework assignments. The number of used late days should be clearly stated in the submitted homework. All late submissions are subject to a penalty of 20% per day for no more than three days.

Exam. There will be take-home midterm and final exams. The midterm will cover material in the previous lectures, and you will be allowed to use one page of note.

Homework and Final Project Policy. Assignments and the final report are due at 11:59 pm Central Standard Time on Friday indicated on the schedule. Students are required to submit the **PDF** file of their project reports. Submissions of homework depend on homework requirements. A word processing software (e.g., LaTeX or Word) is recommended. 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 (4 or more pages). The final project report will be due at the end of 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.