Course Syllabus: COMPSCI 485
Applications of Natural Language Processing
Spring 2024

Instructor: Brendan O'Connor (http://brenocon.com)
TA: Chloe Eggleston (https://chloes.computer/)
Contact: Use a Piazza private message to both instructor and TA. For instructor only, email brenocon@cs.umass.edu with "CS485" in the subject line.

Webpage: https://people.cs.umass.edu/~brenocon/cs485_f23/

Course Description. Natural Language Processing (NLP) is the engineering art and science of how to teach computers to understand human language. NLP is a type of artificial intelligence technology, and it's now ubiquitous -- NLP lets us talk to our phones, use the web to answer questions, map out discussions in books and social media, generate chatbot responses, and translate between human languages. Since language is rich, ambiguous, and very difficult for computers to understand, these systems can sometimes seem like magic -- but these are engineering problems we can tackle with data, math, and insights from linguistics.

This course introduces NLP methods and applications including probabilistic models, text classification, linguistic representations, and neural network language models to process, understand, and generate text. During the course, students will (1) learn the core methods for NLP; (2) become familiar with key facts about human language that motivate them, and help practitioners know what problems are possible to solve; (3) become equipped to engage critically with arguments about NLP's ethical and social implications; and (4) complete a series of hands-on projects to implement, experiment with, and improve NLP models, gaining practical skills for natural language systems engineering.

This course is intended for upper-level undergraduate students in computer science and linguistics, and may be appropriate for interested undergraduate and graduate students from other departments. Its number was previously COMPSCI 490A.

Required Background. This course requires some familiarity with computer programming and probability theory at an intermediate undergraduate level. Formally, the course prerequisite is to fulfill either of:

- Both COMPSCI 220 and COMPSCI 240, with grade "C" or higher; OR
- LINGUIST 492B, with grade "C" or higher

Learning Outcomes. The primary learning outcomes in this course are:

1. Be comfortable with mathematical modeling applied to language, and how it can be implemented in software.
2. Have practice and comfort with technical writing and poster presentation in the context of summarizing research and the student's own experiments.
3. Know basic linguistic concepts and how to manually conduct linguistic analysis, including fundamental challenges in operationalization and construct validity.
4. Be able to make informed arguments about social and ethical issues in NLP.

5. Improve or introduce practical software skills for data analysis, machine learning, and linguistic processing of text.

Course Materials and Technologies: The course will use a freely available textbook and a variety of UMass supported educational technologies.

- **Primary Textbook:** *Speech and Language Processing*, 3rd edition. Dan Jurafsky and James H. Martin. We use a draft version that is freely available online (https://web.stanford.edu/~jurafsky/slp3/) and will supply a copy if necessary. Additional readings will be supplied.

- **Course Website:** Piazza will be the main course website. See link on public course webpage. It will include lecture notes, assignments, readings, demos, and links to online resources.

- **Course Announcements and Discussion Boards:** Official announcements for the course will go out through Piazza. Piazza will also host discussion forums for the class. Students will be added to Piazza by course staff at the start of the semester.

- **Assignment Submission:** Assignments will be submitted using Gradescope. Students will be added to Gradescope by course staff prior to the release of the first assignment.

- **Lecture Videos** will be accessible via Piazza or the website, on Echo360.

Course Topics include text processing, corpus data, probabilistic language models, syntax models, word embeddings, text classification, topic analysis, sentiment analysis, question answering, deep learning methods, and language generation. Topics are subject to change during the semester.

Coursework and Grading Plan. Grades are composed of:

- 10%: **Exercises** (approximately weekly). These are graded with a simplified check/check-minus system. Some are in-class exercises or start as in-class exercises; we will make them available electronically if you miss lecture, but it will be a better educational experience if you attend, and you are more likely to succeed elsewhere. We will drop one exercise from your grade.

- 30%: **Homework assignments (problem sets),** which contain both written and programming portions. We plan to have four homework assignments.

- 30%: **Midterm**, planned to be in-class.

- 30%: **Final project**, including project proposal (10%), progress report (5%), in-class presentation (2.5%), and final report (12.5%). These are done in groups. (All other coursework is independent.)

    o The project presentation will likely be an in-class slide-based presentation. Final plan will be subject to logistical issues and announced later in the semester.

Final grades will be computed using the following grading scale, but we may make adjustments of curving or rescaling when appropriate.

F = lower than 64, D = [64,67), D+ = [67,70), C- = [70,73), C = [74,77), C+ = [77,80), B- = [80,84), B = [84,87), B+ = [87,90); A- = [90,93), A = 93 or higher

Course Policies: Students should make sure they are familiar with all course policies and the relevant University policies linked to below. By staying enrolled in this course, students agree to be bound by all applicable policies.

- **Course Community Code of Conduct:** The instructor and the course staff are committed to providing a friendly, safe and welcoming environment for all, regardless of gender identity and expression, sexual

orientation, disability, personal appearance, body size, race, ethnicity, age, religion, nationality, or other similar characteristic. Please be courteous, respectful, and professional in all of your interactions with other students, TAs, and graders in all mediums of communication including but not limited to in-person, email, video meetings, chat, discussion forums, and re-grade submissions.

Demeaning, insulting or harassing any member of the course community over any medium of communication is not acceptable behavior, including in person, through official course platforms and through personal/private platforms (social media, email, DM, text, etc.). Students who engage in such behavior will be warned at most once before the behavior is reported to the Dean of Students office. If you feel you have been or are being harassed or made uncomfortable by a member of this course community, please contact a member of the course staff immediately (or if you do not feel safe doing so, contact the Chair of the Faculty of CICS, Erik Learned-Miller (chair@cs.umass.edu), or the Dean of Students office). We care about making this course a safe and welcoming place for all.

- **Class Attendance Policy:** Students are encouraged to attend class meetings unless they are feeling ill. Students who are not able to attend class meetings for any reason can stay up to date by watching lecture videos and reading posted course notes, slides, and course announcements. This course does not have a graded attendance component, though in-class exercises are easier to complete in-person.

- **Exam Absence Policy:** A makeup exam time will be provided to students who are unable to attend regular scheduled exams according to University policy (e.g., in the case of illness, religious observances, official University travel, and other extenuating circumstances). Note that a makeup exam will be provided in the case of official University travel (e.g., to present at a conference), but conflicting research deadlines (e.g., a conference paper submission deadline) are not grounds for requesting a makeup exam. When students are aware of an exam conflict ahead of time, they should contact the course staff in writing as soon as possible and no later than one week before the exam date to arrange a time for a makeup exam. In the case of illness or other unforeseen extenuating events, students should contact course staff in writing when they are able. Providing documentation for unforeseen exam absences is greatly appreciated.

- **Late Homework Policy:**
  - For homework assignments:
    - Everyone has **three** late days to use for homework assignments.
    - After all late days have been exhausted, no more non-excused late submissions will be accepted.
  - We follow University policy in accepting work late due to health, personal, and religious holiday circumstances, which are extensions that do not use late days. Arrangements in non-emergency circumstances must be made in *advance* (at least 24 hours) with course staff.
  - Job interviews and other schoolwork are **not** acceptable excuses for late homework.
  - Exercises will be accepted late for the lowest grade (check-minus).
  - Project assignments will not be accepted late.
  - In case of unforseen and emergency circumstances (e.g. illness, internet issues, personal emergency), please contact the instructor when you're able to.

- **Re-grading Policy:** Errors in grading can occur despite the best efforts of the course staff. If you believe you've found a grading error, please submit a re-grade request. Re-grade requests must be submitted no later than one week after the graded material is returned. Note that re-grading may result in your original grade increasing, decreasing or remaining unchanged as appropriate.

- **Academic Honesty and Homework Collaboration Policy:**
  - For all non-group (non-project) assignments, **all of the content you submit, both code and text, needs to be produced independently.** Your work must be in your own words and based on your understanding of the solution. ("Independently" applies to homework; for group projects, work must be produced only by members of the group.)
  - The following are disallowed and extremely likely to result in an academic honesty violations:
    - Sharing completed or in-progress reports or code with another student
    - Copying all or part of completed or in-progress reports or code requested from another student
    - Using another student's code or code output to help debug your code
    - Working with another student while you both write-up or program a solution to a problem
    - Posting completed solutions on public code repositories (e.g., GitHub, etc.) during or after the course
    - Copying any solution materials (derivations, code, method descriptions) in whole or in part from external sources. Disallowed external sources include books, web pages, homework "help" services such as Chegg, automated generation such as ChatGPT, etc.
  - On your homework, list the names of everyone you collaborated or had discussions with.
  - If you find, use, or build off of published material, for example on the web or from a textbook, you must cite the source. Always explain the answer in your own words.
  - For written assignments, all your writing must be original or properly paraphrased, with citations as appropriate.
  - Detected cheating on homework assignments will result in a grade penalty on the assignment up to and including the full value of the assignment. Any misuse of course platforms is grounds for an F in the course. Cheating on exams is grounds for an F in the course.
  - We follow the university's [Academic Honesty Policy and Procedures](). If you are not aware, note that academic honesty is significantly more important than completely finishing assignments or being late. We will report and sanction instances of dishonesty with no leniency. If you have questions about a particular situation, please ask.
  - Some examples of the collaboration policy:
    - Acceptable: Alice and Bob discuss alternatives for storing large, sparse vectors of feature counts, as required by a problem set.
    - Unacceptable: Alice and Bob sit down together and write code for storing feature counts.
    - Acceptable: Bob is confused about how to implement the Viterbi algorithm, and asks Alice for a conceptual description of her strategy.
    - Unacceptable: Alice and Bob divide the assignment into parts, and each write the code for their part, and then share their solutions with each other to complete the assignment.
    - Acceptable: Alice asks Bob if he encountered a failure condition at a "sanity check" in a coding assignment, and Bob explains at a conceptual level how he overcame that failure condition.

- ▪ Unacceptable: Alice or Bob obtain a solution to a previous year's assignment or to a related assignment in another class, and use it as the starting point for their own solution.

- **Course Material Intellectual Property Policy:** The instructor and the University share intellectual property rights for all course materials including lecture slides, lecture audio/video recordings, demo code, assignment handouts, and exam materials. Students are allowed to keep copies of this material for personal use, but are prohibited from distributing it to other individuals and/or posting it in part or in whole on publicly accessible sites including on slide share sites and sites such as Chegg. Students are not permitted to make their own lecture recordings (audio or video). Official recordings will be made available to all students.

- **Course Communication Policy:** Clarification questions about course material or assignments and logistics questions should be submitted to Piazza as public posts. Questions about your solutions to homework assignments and other personal matters should be submitted to Piazza as private posts, which will be viewable by course staff (instructor and graders). For private questions about any personal matters that you do not want to be viewed by any other course staff, please email the instructor, and include "CS485" in your email's subject line. Homework grade requests will be submitted on Gradescope. We aim to answer Piazza posts and emails within 24 hours outside of weekends.

University Course Policy Statements:

- **Accommodation Statement:** The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify the instructor within the first two weeks of the semester so that we may make appropriate arrangements.

- **Academic Honesty Statement:** Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

- **Title IX Statement:** In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need

to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.