

{LING/CS/INFO/ISTA} 539
Statistical Natural Language Processing

Eric Jackson

Last Revised March 8, 2023

1 Course Information

From the course catalog:

This course introduces the key concepts underlying statistical natural language processing. Students will learn a variety of techniques for the computational modeling of natural language, including: n-gram models, smoothing, Hidden Markov models, Bayesian Inference, Expectation Maximization, Viterbi, Inside-Outside Algorithm for Probabilistic Context-Free Grammars, and higher-order language models.

NOTE: The main programming language used in the course will be Python (3.8).

For more information, see the course page in D2L.

Course Objectives

In this course, we will ...

- cover machine learning basics and text classification algorithms, such as ...
 - naive Bayes
 - logistic regression
- explore a range of important natural language processing (NLP) topics, such as ...
 - word representations (ex. embeddings)
 - sequence labeling (part of speech tagging, shallow parsing/chunking, etc.)
 - structured prediction (chart-based parsing, transition-based dependency parsing, etc.)

Learning Outcomes

Students will be able to...

- carry out a variety of natural language processing (NLP) tasks¹
- compare techniques for word and document representations¹
- implement a subset of the algorithms and architectures covered in this class²
- understand an NLP tool or approach well enough to explain it to others.²

Credits: 3 units

Prerequisites

- Programming competency (at the level of ISTA 130 or higher)

Locations and Times

This is an asynchronous online course. According to University of Arizona policy, class attendance is demonstrated by active participation in course-related online activities, such as interacting with D2L pages, the course forum, OpenClass exercises, and so on. Course sessions will not be held in-person. Please see the course D2L page for important dates and further information.

¹Relates to Linguistics Department's UG Program Outcome 1.

²Relates to Linguistics Department's HLT Program Outcomes 1, 2, & 3.

Instructor

Eric Jackson

Email: ejackson1@arizona.edu

Office hours: Thursdays, 10am to 12 noon Arizona time, on Zoom.

<https://arizona.zoom.us/j/89379348121?pwd=QVJPQVdrcTU2NzVucEgxMVptYW1kdz09>

Contact

Students should ask all course-related questions on the ling-539 stream in the course forum (<https://forum.hlt.arizona.edu/#narrow/stream/33-ling-539-sp2023>), not on D2L. If *you* have a question, it's possible that someone else has a similar question. Having the question and answer on the forum means that everyone benefits from it. The course forum is also where you will also find course announcements.

For emergencies or for personal matters that you don't wish to put in a private post, please email the instructor.

Appointments: If you cannot arrange to speak with me during my regular office hours and are unable to discuss course-related issues over email or in a private message on the course forum, you should contact me by email to set up an individual appointment via Zoom.

For planning purposes, please note that I respond to emails and posted questions M–F from 9AM–6PM (MST). Typically, you can expect a response from me within a day.

Schedule

Assessments Please check D2L for assessment details and due dates: <https://d2l.arizona.edu>

Readings & lectures Please check the **Content** section of D2L for unit-specific readings and lectures: <https://d2l.arizona.edu>

Technology

This is a fully online class. As such, you will need a stable internet connection to access course content and submit assignments. To complete your assignments, we recommend that you use a laptop or desktop with ≥ 8 GB of RAM. All assignments and tutorials will be presented using a uniform Linux-based development environment which students will learn to configure during the first week of class (instructions will be provided). To complete your assignments, you will need ...

- A Linux desktop environment such as Ubuntu 22.04 (can be installed as a virtual machine)
- A GitHub account
- git, Docker, and Python 3.8 installed on your linux desktop environment

- A modern web browser (Firefox or Chrome/Chromium)

Virtual office hours will use Zoom.

Python-specific Assistance

For students seeking assistance with Python, the University of Arizona Library's Data Science provides support to students, including through a Slack workspace:

- <https://data.library.arizona.edu/data-science/python-programming>

Other Resources

ResBaz Arizona hosts a variety of events providing researchers opportunities to connect with experienced data scientists and engineers. For more information, see their homepage:

- <https://researchbazaar.arizona.edu/>

Readings

The primary text used in this course is freely available (digital-only):

Jurafsky, Dan and James Martin. Forthcoming. *Speech and language processing*. (3rd ed.)
Draft available at <https://web.stanford.edu/~jurafsky/slp3/>

Required readings (papers, chapters, etc.) will be provided by the instructor.

Supplemental Reading

In addition to the course textbook and any posted readings, students may find the following resources useful:

Goldberg, Yoav. 2017. *Neural Network Methods for Natural Language Processing*. Springer.

Nielsen, Michael. 2015. *Neural Networks and Deep Learning*. Springer. <http://neuralnetworksanddeeplearning.com/>

Bird, Steven, Ewan Klein, and Edward Loper. 2015. *Natural Language Processing with Python*. <https://www.nltk.org/book/>

Géron, Aurélien. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. (2nd ed.) O'Reilly. <http://neuralnetworksanddeeplearning.com/>

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016 *Deep Learning*. MIT press. <https://www.deeplearningbook.org/>

Electronic versions of all the recommended resources listed above are freely to University of Arizona students. Access can be obtained through the UA Library website.

2 Evaluation

	Assessment	Description
65%	Programming Assignments	4-5 assignments that involve implementing algorithms covering in the class. Test cases will be provided to help you refine your solutions. <i>The programming language used will be Python.</i>
25%	Review & Mastery Assignments	Low-risk assessment consisting of guided reviews and questions designed to assist in <i>retaining and mastering</i> material covered in class. We will be using the Open-Class platform for these assignments.
10%	Term project	The term project can be satisfied in one of two ways: :: An original blog post that provides an overview/tutorial to an NLP task/library with an end-to-end walkthrough example. Components include a) a proposal, b) a publicly accessible URL to your blog post, and c) a repository containing the source for the post. :: An entry in a class Kaggle competition involving a realistic classification task. Components include a) a submission to the Kaggle competition itself, b) a blog post documenting the student's approach to the task, and c) a repository containing the source for the Kaggle submission.

Due Dates

Assignment	When?	Where?
Programming Assignments	\geq 1-2 weeks after release	GitHub Classroom (via D2L \rightarrow Assignments)
Review & Mastery Assignments	1-2 weeks after release	OpenClass (via D2L nav bar)
Term project: <i>topic</i>	\approx 4 weeks into course	GitHub Classroom (via D2L \rightarrow Assignments)
Term project: <i>all other components</i>	final day of class*	GitHub Classroom (via D2L \rightarrow Assignments)

Dates above are only estimates and are thus subject to change.

*Final day of class is earlier than the final day of the term; see the course D2L page for an exact date.

3 Grading

Grades will be posted to the course's D2L site.

For assignments involving code, you will be given parameters that describe the behavior of your code. You will be provided with a subset of test cases to help you refine your solution before submitting. Assignment grades will be based on the number of test cases (some of which you have seen on the assignment, and some of which will be secret) your code can pass.

The grading scheme is as follows:

Grade	Point Range
A	90 – 100
B	80 – 89
C	70 – 79
D	60 – 69
E	0 – 59

Grade Disputes

Disputes about grades on a particular project will be entertained for two weeks from the day the project is due, or 1 day before grades are due, whichever is sooner. These will be resolved by re-grading the entire project. Note that this can result in a lower grade in the event that new mistakes are discovered. **No negotiations about individual students' letter grades will be entertained once final grades are assigned, except as permitted by course policy and university policy stated elsewhere in this document.**

Collaboration Policy

Students are encouraged to discuss problems and general approaches for solutions. However, **project implementations and the associated documentation for each project must be completed individually. Copying another person's work (whether that person is also enrolled in this course or not) is not permitted and will be treated according to university policy as a case of academic dishonesty.**

Late Policy

Projects are due electronically via D2L or GitHub Classroom by the stated deadline. Permission for an extension must be granted by the instructor *in advance* of the deadline in order to receive full credit for a late submission. Barring any extraordinary circumstances or accommodations in accordance with university policies, late work is not accepted. No project will be accepted once solutions are posted online.

On Dropping Classes

If you find yourself thinking about dropping this (or any other) class, first make sure that that's what you really want to do. Chatting with the instructor or your academic advisor may help. If you drop within the first week of the term, there will be no record on your transcript; it will be as though you'd never enrolled. After the second week, a drop will be recorded on your transcript. You will receive a "WP" (withdrawn passing) only if you were passing the class at the time of your drop. Toward the end of the term, dropping becomes a challenge, because you need to explain to the instructor and to the dean why you were unable to drop the class during the first half of the term. For drop deadlines specific to this compressed format, please see <https://www.registrar.arizona.edu/dates-and-deadlines>

University boilerplate

All of the following items are required by the university to be included on syllabi. If you find something here that is surprising or unexpected, please bring it up with me as soon as possible.

By way of a brief summary:

Disabilities If you have a disability that affects how you will need to do the work in this class, please let me know *within the first week of class*.

Academic Code of Conduct Cheating and plagiarism are not remotely acceptable in any way. Disruptive behavior in class—which here means on any of our course websites or by email—is not acceptable. Please be respectful of others.

Sensitive Material This is a university and you are adults. It is possible that we may touch on topics that some students could find sensitive during the semester. Given the focus of this course, this seems unlikely, but I alert you nonetheless.

Covid

The university has a specific site for covid information: <http://covid19.arizona.edu>. These are extraordinary times and you may still be experiencing personal and financial challenges. Let me know if we need to make accommodations for covid-related things, and please stay safe.

Absence and Class Participation Policy

Attendance in an all-online course is not evaluated like attendance in an in-person course. For this course, attendance will be represented by active reading, completion, and participation in online course activities, including materials and activities posted on D2L, OpenClass, our course forum, and any other related websites.

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at: <http://catalog.arizona.edu/policy/>

`class-attendance-participation-and-administrative-drop`

The UA policy regarding absences is that any sincerely held religious belief, observance or practice will be accommodated where reasonable, <http://policy.arizona.edu/human-resources/religious-accommodation-policy>.

Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: <https://deanofstudents.arizona.edu/absences>

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities.

Students are asked to refrain from disruptive conversations with others in the course. Students observed engaging in disruptive activity will be asked to cease this behavior. Those who continue to disrupt the class will be asked to leave lecture or discussion and may be reported to the Dean of Students.

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See <http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students>.

Accessibility and Accommodations

At the University of Arizona, we strive to make learning experiences as accessible as possible. If you anticipate or experience barriers based on disability or pregnancy, please contact the Disability Resource Center (520-621-3268, <https://drc.arizona.edu/>) to establish reasonable accommodations.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. If you found a code snippet online, it's important to cite where it came from, even if that source was stackoverflow.com.

Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see

<http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.