Statistical Natural Language Processing
Linguistics 439/539

General

This course focuses on building statistical models of natural language. We do this with *two* aims. First, these models have tremendous value in the practical/computational domain and are widely used in language technology applications. Second, these models have significant appeal as theoretical models of how language is processed or how grammars are organized. We will be using Matlab as our programming environment and programming skills are required for this course. High school algebra will suffice as background, but be forewarned that this course has a significant mathematical component. Linguistics 438/538 is also a prerequisite.

Instructor

Name:	Mike Hammond
Email:	hammond **at** u **dot** arizona **dot** edu (reliable)
Phone	621-5759 (unreliable)
Office:	Douglass 308
Hours:	Wednesdays 9:30–11:30, and by appointment

Text

Manning, Christopher D. & Hinrich Schütze. 1999. *Foundations of statistical natural language processing.* Cambridge: MIT Press.

Course Website

`http://dingo.sbs.arizona.edu/~hammond/ling439-f14/`

Requirements

Homework	$15\% \times 5 = 75\%$
Final homework	25%

There are six regular homework assignments over the term and a final assignment; the lowest *regular* homework score—or missing score—will be dropped. All assignments have additional separate components for students enrolled in 539.

Schedule

Week	Date	Topic	Reading	Due
1	8/26	Paperwork		
	8/28	Overview	MS ch.1	
2	9/2	Matlab		
	9/4			
3	9/9	Probability	MS chs.2–4	
	9/11			
4	9/16	Information Theory		HW#1
	9/18			
5	9/23	N-gram models	MS chs.5–8	
	9/25			
6	9/30	N-gram models continued		HW#2
	10/2			
7	10/7	HMMs	MS chs.9–10	
	10/9			
8	10/14	HMMs continued		HW#3
	10/16			
9	10/21	PCFGs	MS chs.11–12	
	10/23			
	10/28	PCFGs continued		HW#4
	10/30			
10	11/4	Alignment/MT	MS ch.13	
	11/6			
11	11/11	**no class**		[HW#5]
	11/13	Clustering	MS ch.14	
12	11/18			
	11/20	Categorization	MS ch.16	
13	11/25		TBA	HW#6
	11/27	**no class**		
14	12/2	Search	TBA	
	12/4			
15	12/9	TBA	TBA	
	12/11	**no class**		
16	12/16	**no class**		final HW

Readings are to be done *before* class on the day of the week for which they are listed. Be aware that due dates are real. Programming components of assignments must be emailed to me as `m-files`; prose components may be emailed as pdf (*not* MSWord) or turned in as hardcopy. In all cases, assignments are due by the beginning of class on the day indicated. *Late work will not be accepted.*

Software

We will be making extensive use of Matlab. *Please download, install, and confirm that it's working by the first week of classes.*

- `http://softwarelicense.arizona.edu/mathworks-matlab`
  (glitzy and expensive, but *free* for UA folks)

It is your responsibility to make sure any programs you write run on the latest version of Matlab regardless of operating system. In other words, do *not* use OS-specific code and do not use earlier versions of Matlab. (Current version: `R2014a`)

Et cetera

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 acceptable. Disruptive behavior in class is not acceptable.

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.