1. Description and Goals

Computers have made it possible, even easy, to collect vast amounts of text from a wide variety of sources. It is not always clear, however, how to use those data and how to extract useful information from data. This problem is faced in a tremendous range of scholarly, government, business, medical, and scientific applications. The purpose of this course is to teach some of the best and most general approaches to get the most out of natural language.

Finally, it is required that you have regular access to a computer and an Internet connection throughout this course. A laptop is preferable. If you have a laptop, it would be useful to bring the laptop to class, especially for the lab sessions (see below).

2. Approach

This is a flipped course. Lectures are delivered through the Internet, and the traditional “class time” is used for hands on projects, discussion, and working on homework. To make sure that you are watching the videos, there will be a quiz at the start of every class on the material covered in the videos.

Halfway through the course, students will complete an in-class midterm that will test high-level understanding of concepts.

In addition, students will work on a project that emphasizes the concepts covered in the class. The project will be in a group of three to four students focusing on a shared problem. This project will have three stages:
an initial proposal, explaining what they want to do
• a first-step deliverable (due three quarters into the semester) to make sure students are grappling with
  the problem, e.g. a transformation of the data, performing a preprocessing step, or creating a baseline for
  comparison
• a final project report

2.1 Preliminary Schedule (Check webpage for current version)

1. Morphology
2. Language Models
3. Part of Speech Tagging
4. Classification
5. Hidden Markov Models
6. Constituency Parsing
7. Dependency Parsing
8. Anaphora / Coreference
9. Machine Translation
10. Topic Models
11. Semantics

2.2 Required Background

Mathematical maturity: We will work extensively with probability and mathematical functions such as logarithms and
differentiation. You should be comfortable manipulating these concepts algebraically. You should also be able to argue
why mathematical statements are true (even if it’s not a 100% formal proof).

We will make extensive use of the Python programming language. It is assumed that you know or will quickly
learn how the program in Python. Apart from a quick introduction, there will be no introduction to this skill-set.

The computer-based aspects of this course will be oriented toward Unix-like operating systems (Linux, OS X). It may be
possible to complete the course using other operating systems, but you will be responsible for troubleshooting any issues
you encounter.

3. Grading

Components of the final grade are as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>30%</td>
</tr>
<tr>
<td>Exams</td>
<td>25%</td>
</tr>
<tr>
<td>Final Project</td>
<td>25%</td>
</tr>
<tr>
<td>Quizzes</td>
<td>10%</td>
</tr>
<tr>
<td>Participation</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

It is possible to earn extra credit by going above and beyond the expectations of the assignment or by attending
relevant extra-curricular events.

Letter grades will be assigned as follows:

https://docs.google.com/document/u/1/d/1nTkylpljzNs0ORk7GXbN2ez4X0eoIT5v655gW2RmgM/pub
9/12/2018  CL1 Syllabus

- 97.00+    A+
- 93.00-96.99    A
- 90.00-92.99    A-
- 87.00-89.99    B+
- 83.00-86.99    B
- 80.00-82.99    B-
- 77.00-79.99    C+
- 73.00-76.99    C
- 70.00-72.99    C-
- 60.00-69.99    D
- 0.00-59.99    F

I reserve the right to curve up the threshold (i.e. a lower point value may result in a higher grade), but I will not curve down (i.e., a higher point value will not result in a lower grade). The thresholds will be placed uniformly for the entire class.

Please note that if the final grade tabulation comes out to be 79.98, then that corresponds to a C+; I have been exact in the above specifications deliberately. I’m sorry, but if I negotiate on any of these cutoffs, I then need to negotiate on the next one (e.g. if I rounded 79.95 up, then I would get harrased about 79.94). Especially for large classes, this results in chaos. I need to draw the line somewhere, and I’m sorry if fall on the wrong side of the line (I’ve had that happen to me before too).

3.1 Assignments

There will be a number homework assignments (with different numbers of points). Together, they are worth 30% of your final grade. Assignments are designed to help you learn the material, so please use them for that! You are allowed to collaborate with others (as many people as you'd like), but you must turn in your own assignment. For example, you could work together in a group, but each person must write up their solutions individually. Everything you submit must be your own. Directly from your fingers. Copying and pasting from another student will be considered plagiarism and will be handled according to the university's academic integrity policies.

Assignments are due at 11:59 Eastern on the Friday indicated on the syllabus. Late policy: each person has seven free late days to be used, no questions asked, during the course. Late assignments beyond the provided late grades will not be graded. The grade of the lowest homework assignment will be thrown out.

3.2 Exams

There will be an in-class midterm and final. The midterm will cover material in the previous lectures and you will be allowed to use one page of handwritten notes and the course text. Likewise, the final will be cumulative but biased toward end of course.

3.3 Final Project

More information will be posted on a separate page for the final project.

3.4 Class Participation

Each class is critical to your learning experience, and I expect you to come to class prepared (having read all assigned readings, ready to engage). I also expect active participation, not passive reception of the material. Your energy in contributing to class discussions and hands-on exercises will make this class an enjoyable experience for all of us. Class participation contributes to your total grade (see above).
We will also be using the online learning platform Piazza. You can get credit for participation by answering and asking useful questions on that platform. Ideally you should be participating both online and in class, however.

4. Academic Integrity

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit http://www.shc.umd.edu.

5. Course Policies
The University has a legal obligation to provide appropriate accommodations for students with disabilities. Please inform the professor of any accommodations needed relative to disabilities at the start of the semester.

Also, University of Maryland policy states that students should not be penalized due to observances of their religious beliefs. Please inform the professor of such instances at the start of the semester so that appropriate steps can be taken.
Computational Linguistics I (INST 735 / CMSC 723 / LING 723)

Logistics

Location  CSIC 3117
Time       Mon./Wed. 15:30pm - 16:45pm
Webpage    http://umiacs.umd.edu/~jbg/teaching/CMSC_723/
Mailing List https://piazza.com/umd/fall2018/cmsc723
Text       Natural Language Processing
Syllabus   https://docs.google.com/document/d/1nTkyPlijzNs0ORk7GXbN2ec4X0eoIT5v655gW2RawgM/pub

People

Professor

Jordan Boyd-Graber
AVW 3153
Office Hours (AVW 3155): Starting Sept. 4, Mondays 14:00 - 15:00 and by appointment

Teaching Assistants

Chen Zhao: AVW 4424, Thursdays 15:00-16:00 Ahmed Elgohary: AVW 4185, Mondays 13:00-14:00

Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>In-Class Topic</th>
<th>Assignment Due</th>
<th>Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon 27. Aug</td>
<td>Introduction to the course,</td>
<td>[Intro Video]</td>
<td></td>
</tr>
</tbody>
</table>
# Probability, and Python

## Optional Readings:

- Python (and programming) [introductory course](#)
- You may also want to look at first couple of weeks (through February) of [INST 414](#) as review.
- [A probability primer](#)
- (Optional, if you want more detail) [Grinstead and Snell, Chapters 1-2, Chapter 4.1](#)
- How the statistical revolution changes (computational) linguistics
- On the [deep learning revolution in NLP](#)

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wed 29. Aug</td>
<td>Probabilistic Classification (Ahmed)</td>
<td>[Slides: NB, LR, Ex] [Video: NB, LR]</td>
</tr>
<tr>
<td>Mon 3. Sep</td>
<td>No Class: Labor Day!</td>
<td></td>
</tr>
<tr>
<td>Wed 5. Sep</td>
<td>Classification II</td>
<td>[Slides: SG Perceptron SVM, Ex] [Video: SGD Support Vector Machines, Perceptron]</td>
</tr>
<tr>
<td>Fri 7. Sept</td>
<td>HW 0</td>
<td>Limericks</td>
</tr>
<tr>
<td>Mon 10. Sep</td>
<td>Deep Learning</td>
<td>[Slides: Class, Deep, Ex] [Video: Deep Backprop]</td>
</tr>
<tr>
<td>Wed 12. Sep</td>
<td>Course Project (Chen)</td>
<td></td>
</tr>
<tr>
<td>Mon 17. Sep</td>
<td>Distributional Semantics</td>
<td>[Video: Intuition, Algorithm, Evaluation]</td>
</tr>
<tr>
<td>Wed 19. Sep</td>
<td>Frameworks</td>
<td>[Video: Intro Pytorch DAN, Slides: Classification, Compgraph, Code, DAN, Ex]</td>
</tr>
</tbody>
</table>

[http://users.umiacs.umd.edu/~jbg/teaching/CMSC_723/](http://users.umiacs.umd.edu/~jbg/teaching/CMSC_723/)
Fri 21. Sept  HW 1 Classification
Mon 24. Sep Language Models [Video: Intro Backoff]
Readings:
  * NLP 6

Wed 26. Sep Backoff and Smoothing
Readings:
  * Chen and Goodman

Optional:
  * Dan Jurafsky on KN
  * Mike Collins on Discounting
  * Bayesian Interpretation of Interpolated Kneser-Ney

Fri 28. Sept  HW 2 Language Models
Mon 1. Oct Neural Sequence Models [RNN] [LSTM]
Readings:
  * NLP Chapter 4
  * Chapter 1 of Applications of Topic Models

Fri 5. Oct  HW 3 Deep Learning
Mon 8. Oct Part of Speech Tagging
Readings:
  * NLP 7-7.4
  * Unsupervised HMMs

Wed 10. Oct Named Entities and Coreference
Readings:
  * NLP 8, 15

Mon 15. Oct FSTs and Morphology [Morphology]
Readings:
  * NLP 9

Wed 17. Oct Constituency Grammars [Video]

Fri 5. Oct  HW 4 Topic Models
Readings:
  * NLP 10
<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Readings</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon 22. Oct</td>
<td>Dependency Grammars</td>
<td>Readings:</td>
<td>Video</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NLP 11</td>
<td></td>
</tr>
<tr>
<td>Wed 24. Oct</td>
<td>Midterm Review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri 26. Oct</td>
<td>Project Proposal Due</td>
<td>Proposal</td>
<td></td>
</tr>
<tr>
<td>Mon 29. Oct</td>
<td>Midterm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed 31. Oct</td>
<td>TBD (Slack Day)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri 5. Oct</td>
<td>HW 5</td>
<td>Parsing</td>
<td>Video</td>
</tr>
<tr>
<td>Mon 5. Nov</td>
<td>Machine Translation</td>
<td>Readings:</td>
<td>Video</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• NLP 18</td>
<td></td>
</tr>
<tr>
<td>Wed 7. Nov</td>
<td>No Class (EMNLP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri 9. Oct</td>
<td>HW 6</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>Mon 12. Nov</td>
<td>Neural MT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed 14. Nov</td>
<td>RL for NLP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri 16. Nov</td>
<td>HW 7</td>
<td>MT</td>
<td></td>
</tr>
<tr>
<td>Mon 19. Nov</td>
<td>Computational Social Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mon 26. Nov</td>
<td>Reading and Reviewing NLP Papers</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Seven Deadly Sins of AI Predictions</td>
<td></td>
</tr>
<tr>
<td>Mon 3. Dec</td>
<td>Project Workshop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wed 5. Dec</td>
<td>Final Project Presentations I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mon 10. Dec</td>
<td>Final Project Presentations II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fri 14. Dec</td>
<td>Final Exam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Page by Jordan Boyd-Graber.