Overview

The rapid proliferation of the internet, mobile devices, social media and related technologies in daily life has presented a profound opportunity for social scientists to pose and test theories of social phenomena in novel advantageous ways. Our lives are now conducted online in such scope and variety, from social interactions to political blogging, topical web searching and consumption of digitally archived and accessed media, that online digital artifacts are beginning to be probed to reveal patterns of social organization and behavior in situ and at scales impossible to achieve using traditional social scientific methodology. A new cross-disciplinary field has emerged, Computational Social Science (CSS), with the aim of exploiting this opportunity. This course provides an overview of the current state of CSS and a practical, hands-on introduction to some of the most common tools employed in CSS. It is intended that students will complete the course with sufficient knowledge and experience to use existing tools to conduct CSS research, and with sufficient preparation for further formal or self-directed study into methods development and evaluation.

The course consists of three components: (1.) In class presentations and discussions of the most current CSS literature covering methods, applications, validation, and philosophical issues; (2.) Lectures covering background material and the most common machine learning tools applicable to social science research; (3.) Homework laboratories exercising the methods using standard text corpora from the machine learning community.

Lecture and Laboratories Outline

<table>
<thead>
<tr>
<th>Week</th>
<th>Lecture Subject</th>
<th>Lab Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/2</td>
<td>Overview and Intro to CSS</td>
<td>Lab 0: Verify Course Account</td>
</tr>
<tr>
<td>4/8</td>
<td>Intro to Python</td>
<td>Lab 1: Python Basics</td>
</tr>
<tr>
<td>4/16</td>
<td>Overview of Probability</td>
<td>None</td>
</tr>
<tr>
<td>4/23</td>
<td>Naive Bayes Models</td>
<td>Lab 2: Naive Bayes Classification</td>
</tr>
<tr>
<td>4/30</td>
<td>Nearest Neighbor Models</td>
<td>Lab 3: kNN Classification</td>
</tr>
<tr>
<td>5/7</td>
<td>Support Vector Machines</td>
<td>Lab 4: SVM Classification</td>
</tr>
<tr>
<td>5/14</td>
<td>k-Means, Spherical k-Means</td>
<td>Lab 5: Spherical k-Means Clustering</td>
</tr>
<tr>
<td>5/21</td>
<td>Hierarchical and/or Spectral Clustering</td>
<td>Lab 6: Clustering 2</td>
</tr>
<tr>
<td>5/28</td>
<td>Non-Negative Matrix Factorization</td>
<td>Lab 7. NMF Topic Identification and Clustering</td>
</tr>
<tr>
<td>6/4</td>
<td>TBD</td>
<td>None</td>
</tr>
</tbody>
</table>

Possible Items for TBD Lecture(s)

- Student Interest.
- Ideological Scaling.
- EM Algorithm in General and on the Unit Hypersphere.
- Twitter API Data Harvesting.
- Analysis of Student Datasets.
- Databases Setup and Use for Managing CSS Datasets.
- Visualization Basics with matplotlib.
- Bag of Words Processing and NLP Basics in Python with NLTK.

Course Website

We will post lecture slides, readings and student presentations on the course website as we go. See below for details.

Grading

* 50% Class participation and reading presentations.
* 50% Lab reports and in class discussion.

Office and Lab Hours:

TA/Computer Lab: CSB 115, Mondays and Fridays 9AM-1PM (tentative).
Rona-Tas SSB 488, TTh 11-11:50AM.
Hunter: Wednesday after class or by appointment.

The Computer Lab

Location: CSB 115 (Cognitive Science Building).
Access Code: TBD

Enrolled students have all been assigned a course account (e.g. username like so290s**) that you can log into using your UCSD email account as password.

Account lookup tool: https://sdacs.ucsd.edu/~icc/index.php
Password setup page: http://acms.ucsd.edu/students/gpasswd.html
Host Machines: so290s**@ieng6.ucsd.edu (standard server does not allow UI access remotely)
so290s**@ieng6-240 through so290s**@ieng6-254 (use these remotely for UI access with VCNgnome)
Remote Access UI: http://acms.ucsd.edu/info/vncgnome.html

NOTE: Do not leave processing running on these machines as they will be killed by the administrator. The labs can be completed in the time of a normal interactive session (maybe with a coffee break to wait for something in some cases).

ACMS Contact Info for Account Troubleshooting:

Maryam Sarkhosh
maryam@acsmail.ucsd.edu
Help Desk: (858) 534-ACMS (2267)

Laboratories

Labs will be done in Python using template programs and lab instructions we provide. The labs are designed for minimal programming background and provide simple step-by-step instructions for using Python and scikit-learn to conduct CSS experiments.

**Reports as PDFs are due as uploads to the course website the day prior to the following class meeting** Name the pdf file soc211_lab##_lastname_firstname.pdf. There are no reports for Lab 0 (setting up and testing account) and Lab 1 (basic Python exercises). Discuss account and tool troubleshooting with the TA in lab hours, and we will discuss basic python exercises in class.

Readings and Presentations

Students will work together in 8 groups of 3-5. Each meeting, 2-3 groups will be responsible to present one of the readings with a short powerpoint slideshow. It is up to student groups to work together outside of class to discuss the readings, collect their summary and analysis of papers and construct the slides. Groups can decide who will responsible for presenting. Since there are 2-3 presentations per group, different students can present different papers. In a few cases, papers are longer and presentation can be split if that is preferred (but please construct and submit only one presentation pdf).
Slide presentations as PDFs are due as uploads to the course website the day prior to the date of presentation. Name the pdf file soc211_group##_leadauthor_year.pdf.

Readings Presentations by Week:

<table>
<thead>
<tr>
<th>Date/Group</th>
<th>Theme/Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/8</td>
<td>Big Data Issues, Web Science</td>
</tr>
<tr>
<td>4/16</td>
<td>Text Analysis, Machine Learning</td>
</tr>
<tr>
<td>4</td>
<td>Domingos, P. (2012)</td>
</tr>
<tr>
<td>4/23</td>
<td>Measuring Culture, Political Movements, Language Dynamics</td>
</tr>
<tr>
<td>6</td>
<td>Bail, C. A. (2014)</td>
</tr>
<tr>
<td>7</td>
<td>Hanna, A. (2013)</td>
</tr>
<tr>
<td>4/30</td>
<td>Sentiment, Implied Support and Opposition</td>
</tr>
<tr>
<td>5/7</td>
<td>Classification of Political Language</td>
</tr>
<tr>
<td>5/14</td>
<td>Clustering</td>
</tr>
<tr>
<td>7</td>
<td>Hopkins, D. J., &amp; King, G. (2010)</td>
</tr>
<tr>
<td>5/21</td>
<td>Topic Models, Latent Variables</td>
</tr>
<tr>
<td>8</td>
<td>Blei, D. M. (2012)</td>
</tr>
<tr>
<td>5/28</td>
<td>Ideological Scaling</td>
</tr>
<tr>
<td>TBD</td>
<td>Laver, M., Benoit, K., &amp; Garry, J. (2003)</td>
</tr>
</tbody>
</table>

Required Readings for Presentation


Supplemental Readings