Course Description
In the past several years, social scientists have been facing a quantitative change in technology. This change can be summarized in two main points: 1. availability of vast and seemingly insurmountable volumes of human-related data, and 2. constantly increasing computational power. These have provided an unprecedented opportunity to study and model human cognition with range and detail previously not imaginable. Moreover, there is growing interest (e.g. in marketing) to use such data for predicting a variety of human behavior, for detecting different types of activities, or for more intelligent targeted advertising. Advanced Big Data Methods I focuses on methods in computer science, specifically in machine learning, which can help us achieve these outcomes. This course is followed by Advanced Big Data Methods II which focuses on the applications of machine learning techniques, along with natural language processing and network analysis, guided by psychological theories, for identifying various social and cognitive properties evident in human related big data.

The intended audience for this course is psychology graduate students, and more broadly graduate students in social sciences, who are interested in using machine learning techniques for analysis of data. Also, this course may be of interest to PhD students in communications, computer science and the business school.

Learning Objectives
This course is designed to be hands-on and students are expected to learn how to apply different machine learning techniques for analyzing different types of data. In order to achieve this objective, each discussed topic is accompanied by a lab session in which we examine how to use that technique on a data set. Lab clinic sessions are used for helping students troubleshoot their code and also for going over the homework.

Prerequisite(s): Instructor permission
Recommended Preparation: Psych 501 or a similar introductory statistics course

Course Notes
Lecture notes and homework assignments will be posted on Blackboard. Students are also highly encouraged to use the course forum on Blackboard.

Technological Proficiency and Hardware/Software Required
This class includes lab sessions. Students are required to bring a laptop to class. Homework assignments are programming problems that need to written in R.
Required Readings and Supplementary Materials

Required:
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with Applications in R. Springer.

Supplementary:

Description and Assessment of Assignments

a. Homework assignments. Each week students will complete programming problems from the required book. The assignments will be graded based on both output and style of the code. The homework material will be reviewed during lab clinics.
b. Lab presentation. Each student will do a lab presentation in which a particular lab module is taught to others. Also, it is recommended that the presenter extends the lab module by applying the method being discussed to some other data sets.
c. Class Project. Students need to do a project for this class. The goal of the project is for students to get experience in applying big-data methods for analyzing behavioral data. This will include a project proposal presentation, final project presentation, and a report. For project proposals, students will present a problem and a data collection method and/or dataset for which they want to analyze using the methods discussed in class. Each presentation should be about 10-15mins. For the final project presentation, each student/group will give a 15-20min presentation on their project. Students are expected to spend at least 60 hours working on their final project. The project report will be 10-15 pages.

Grading Breakdown

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Points</th>
<th>% of Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>Lab presentation</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Project Proposal</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Project Presentation</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td>Final Project Report</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>Participation</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>TOTAL</td>
<td>530</td>
<td>100</td>
</tr>
</tbody>
</table>

Assignment Submission Policy

Homework will be assigned on Thursdays and will be due the following Thursday at 11am, before the start of class submitted on Blackboard. All homework turned in any later than 11:10am will be considered late. Students will be allowed a total of seven late days that can be used on the assignments. In exceptional circumstances, arrangements must be made in advance of the due date to obtain an extension. Once you have used up your seven late days, one additional day late will result in a 25% reduction in the total score, two additional days late will yield a 50% reduction, and no credit will be given for three or more additional days late. Late days are in units of days, not hours, so using up part of a day uses up the whole day. The final project report, plus the R code used, will be due on the day of the final exam.
Course Schedule:
The following schedule is tentative and may change during the semester.

1. **Week 1 January 15th**: Introduction, Statistical Learning & Linear Regression  
   a. What is Statistical Learning (ISLR 2.1)  
   b. Assessing Model Accuracy (ISLR 2.2)  
   c. Simple & Multiple Linear Regression (ISLR 3.1 & 3.2)  
   d. Overview of Classification (ISLR 4.1)  
   e. Logistic Regression (ISLR 4.3)  
   f. HW 1 assigned

2. **Week 2 January 22**: Classification Cont., Lab 1 & 2, Lab Clinic 1  
   a. Linear Discriminant Analysis (ISLR 4.4)  
   b. Comparison of Classification Methods (ISLR 4.5)  
   c. Bayesian Classifiers  
   d. Lab Clinic 1  
   e. Lab: Linear Regression (ISLR 3.6)  
   f. Lab: Logistic Regression, LDA, QDA and KNN (ISLR 4.5)  
   g. HW 1 due, HW 2 & 3 assigned

3. **Week 3 January 29**: Resampling Methods, Support Vector Machines, Lab Clinic 2  
   a. Cross Validation (ISLR 5.1)  
   b. The Boot Strap (ISLR 5.2)  
   c. Maximal Margin Classifier (ISLR 9.1)  
   d. Support Vector Classifiers (ISLR 9.2)  
   e. Support Vector Machines (ISLR 9.3 & 9.4)  
   f. Support Vector Regression (Handouts)  
   g. Lab Clinic 2  
   h. HW 2 due, HW 4 assigned

4. **Week 4 February 5**: Tree-Based Methods, Lab 3 & 4, Lab Clinic 3  
   a. Decision Trees (ISLR 8.1)  
   b. Bagging, Random Forests, Boosting (ISLR 8.2)  
   c. Lab: Resampling Methods (ISLR 5.3)  
   d. Lab: Support Vector Machines (ISLR 9.6)  
   e. Lab Clinic 3  
   f. HW 3 due, HW 5 assigned

5. **Week 5 February 12**: Unsupervised Learning, Lab 5-6, Lab Clinic 4  
   a. Principle Component Analysis (ISLR 10.2)  
   b. Clustering Methods (ISLR 10.3)  
   c. Other Methods (Handouts)  
   d. Lab: Decision Trees (ISLR 8.3)  
   e. Lab: Unsupervised Learning (ISLR 10.4 – 10.6)  
   f. Lab Clinic 4  
   g. HW 4 due, HW 6 assigned

6. **Week 6 February 19**: Project Proposal Presentation & Review of probability calculus, Lab Clinic 5  
   a. Project Proposal Presentation  
   b. Lab Clinic 5  
   c. HW 5 due

7. **Week 7 February 26**: Bayesian networks: syntax and semantics, Lab Clinic 6  
   a. From Propositional to Graded Beliefs (Handouts)  
   b. Updating Beliefs (Handouts)  
   c. Independence (Handouts)  
   d. Capturing Independence Graphically (Handouts)  
   e. Parameterizing the Independence Structure (Handouts)  
   f. Reasoning with Bayesian Networks (Handouts)
8. Week 8 March 5: Modeling with Bayesian Networks, Lab 7
   a. Handouts
   b. Lab Clinic 6
   c. HW 6 due
   d. Lab 7: Bayesian Network in R
   e. HW 7 assigned
9. Week 9 March 12: Bayesian networks: structure learning & Inference, Lab Clinic 7
   a. Constraint-Based Structure Learning Algorithms (BNR 2.2.1)
   b. Score-Based Structure Learning Algorithms (BNR 2.2.2)
   c. Inference in Bayesian Networks (Handouts)
   d. Lab Clinic 7
   e. HW 7 due
10. Week 10 March 19: Spring Break
11. Week 11 March 26: Introduction to Search and Network Analysis & Lab 8
   a. Handouts
   b. Lab: Structure Learning with Bayesian Networks in R (BNR 2.3)
   c. HW 8 assigned
12. Week 12 April 2: Introduction to text analysis & Lab Clinic 8
   a. Introduction to text analysis (Handouts)
   b. Lab Clinic 8
   c. HW 8 Due
13. Week 13 April 9: Deep Learning I & Lab 9
   a. Artificial Neural Networks
   b. Papers
   c. Lab: iGraph package
   d. HW 9 assigned
14. Week 14 Apr 16: Deep Learning II, Lab 10, Lab Clinic 9
   a. Papers
   b. Lab: Text analysis in R
   c. HW 10 assigned
   d. Lab Clinic 9 – HW 9 due
15. Week 15 Apr 23: Lab Clinic 10
   a. HW 10 due
16. Week 16 April 30: Final Project Presentations
17. May 7: Final Assignment and Project Reports Due

Statement on Academic Conduct and Support Systems

Academic Conduct
Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Section 11, Behavior Violating University Standards https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity http://equity.usc.edu/ or to the Department of Public Safety http://capsnet.usc.edu/department/department-public-
safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men [http://www.usc.edu/student-affairs/cwm/] provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems
A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute [http://dornsife.usc.edu/ali], which sponsors courses and workshops specifically for international graduate students. The Office of Disability Services and Programs [http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html] provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, USC Emergency Information [http://emergency.usc.edu/] will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.