Psych 625: Applied Machine Learning

Morteza Dehghani

Fall 2020

E-mail: mdehghan@usc.edu
Office Hours: Tues 10-12
Office: SGM 607

Web: cssl.usc.edu
Class Hours: Th 2-6pm
Class Room: ONLINE

Course Description

In the past decade, 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. Applied Machine Learning focuses on methods in computer science, specifically in machine learning, which can help us achieve these outcomes. This course is followed by Psych 626: *Text as Data* which focuses on the applications natural language processing, 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 be written in *R*.

Required Books

- Gareth, J., Daniela, W., Trevor, H., and Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger

- Chollet, F. and Allaire, J. J. (2018). *Deep Learning with R*. Manning Publications Company

Description and Assessment of Assignments

1. Homework assignments. Each week students will complete programming problems from one of the required books. The assignments will be graded based on both output and style of the code. The homework material will be reviewed during lab clinics.

2. Lab presentation. Each student will do a lab presentation in which a particular lab module is taught to others using a different dataset than the one used in the book. This dataset will be made available to the class by the second week of class. By having enrolled in the class all student acknowledge the copyright information regarding this dataset.

3. Class Projects. Students will complete four class projects. These projects will be relatively heavy programming assignments requiring students to use R to implement some specific statistical technique. The first project will be relatively easy and not time consuming. The other projects, however, will take substantial time (40 hours each). The projects will be assigned at the beginning of the semester.

Grading Policy

- **20%** Homework
- **5%** Lab Presentations
- **5%** Project 1
- **20%** Project 2
- **20%** Project 3
- **20%** Project 4
- **10%** Project 5

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. Usually, three questions will be assigned for homework, and you have the option of answering two of them. 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. All assignments, including the projects, need to be written using *knitR*. Copied and pasted code/results will not be accepted.

Schedule and weekly learning goals

The schedule is tentative and subject to change.

Week 01, 08/20: : Introduction, Statistical Learning & Linear Regression

- What is Statistical Learning (ISLR 2.1)

- Assessing Model Accuracy (ISLR 2.2)

- Simple & Multiple Linear Regression (ISLR 3.1, 3.2 & 3.3)

- Intro to *knitR* package

- HW 1, Project 1 assigned

Week 02, 08/27: Classification, Lab 1, Lab Clinic 1

- Overview of Classification (ISLR 4.1, 4.2)

- Logistic Regression (ISLR 4.3)

- Linear Discriminant Analysis (ISLR 4.4)

- Comparison of Classification Methods (ISLR 4.5)

- Bayesian Classifiers

- Lab Clinic 1

- Lab: Linear Regression (ISLR 3.6)

- HW 1 due, HW 2 assigned

Week 03, 09/03:   Resampling Methods, Linear Model Selection, Lab 2, Lab Clinic 2

- Cross Validation (ISLR 5.1)

- The Boot Strap (ISLR 5.2)

- Subset Selection (ISLR 6.1)

- Shrinkage Methods (ISLR 6.2)

- Dimension Reduction Methods (ISLR 6.3)

- Considerations in High Dimensions (ISLR 6.4)

- Lab Clinic 2

- Lab: Logistic Regression, LDA, QDA and KNN (ISLR 4.6)

- HW 2 due, HW 3 & 4, Project 2 assigned

- Project 1 due

Week 04, 09/10:   Moving Beyond Linearity, Lab 3 & 4

- Polynomial Regression (ISLR 7.1)

- Step Functions (ISLR 7.2)

- Basis Functions (ISLR 7.3)

- Regression Splines (ISLR 7.4)

- Smoothing Splines (ISLR 7.5)

- Local Regression (ISLR 7.6)

- Generalized Additive Models (ISLR 7.7)

- Lab: Resampling Methods (ISLR 5.3)

- Lab: Regularization (ISLR 6.5, 6.6 & 6.7)

- HW 4 due, HW 5 assigned

Week 05, 09/17:   Tree-Based Methods, Support Vector Machines, Lab 5, Lab Clinic 3

- Decision Trees (ISLR 8.1)

- Bagging, Random Forests, Boosting (ISLR 8.2)

- Maximal Margin Classifier (ISLR 9.1)

- Support Vector Classifiers (ISLR 9.2)

- Support Vector Machines (ISLR 9.3 & 9.4)

- Support Vector Regression (Handouts)

- Lab: Moving Beyond Linearity (ISLR 7.8)

- Lab Clinic 3

- HW 5 due, HW 6, 7 assigned

Week 06, 09/24:   Unsupervised Learning, Lab 6 & 7, Lab Clinic 4

- Principle Component Analysis (ISLR 10.2)

- Clustering Methods (ISLR 10.3)

- Lab: Decision Trees (ISLR 8.3)

- Lab: Support Vector Machines (ISLR 9.6)

- Lab Clinic 4

- HW 6,7 due, 8 assigned

Week 07, 10/01:   Recap ISLR

- Project 2 due

- Project 3 assigned

- Lab Clinic 5

Week 08, 10/08:   Neural Networks 1/4

- The Perceptron (Handouts)

- Multilayer Perceptron (Handouts)

- Backpropagation Algorithm (Handouts)

- HW 8 due

- Lab Clinic 6

Week 09, 10/15:   Neural Networks 2/4

- Deep Learning in R (DLR) Chapters 1-4

- Introduction to *keras*

Week 10, 10/22:   Neural Networks 3/4

- Deep learning for text and sequences (DLR Chapter 6)

- Project 3 due

- Project 4 & 5 assigned

Week 11, 10/29:   Neural Networks 4/4

- Deep learning best practices (DLR Chapter 7)

- Generative deep learning (DLR Chapter 8)

Week 12, 11/05:   Deep Learning in the Social Sciences

- Papers TBD

Week 13, 11/12:   Project 4 due

Week 14, 11/19:   Project 5 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.

References

Chollet, F. and Allaire, J. J. (2018). *Deep Learning with R*. Manning Publications Company.

Gareth, J., Daniela, W., Trevor, H., and Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.