Psych 625: Applied Machine Learning

Morteza Dehghani

Fall 2022

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

Web: morteza-dehghani.net
Class Hours: Th 2-6pm
Class Room: VHE210

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. Homework review 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. (2021). *An introduction to statistical learning: with applications in R (Second Edition)*. Spinger

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 homework review sessions.

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 five 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.

Grading Policy

- **25%** Homework
- **5%** Project 1
- **20%** Project 2
- **15%** Project 3
- **10%** Project 4
- **15%** Project 5
- **5%** Presentations
- **5%** Participation

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, 08/25: : Introduction, Statistical Learning & Linear Regression

- ISLR Chapter 2
- ISLR Chapter 3
- Lab: Linear Regression (ISLR 3.6)
- Intro to *knitR* package
- HW 1; Project 1 assigned

Week 02, 09/01: Classification

- ISLR Chapter 4
- Bayesian Classifiers
- Lab: Logistic Regression, LDA, QDA and KNN (ISLR 4.6)
- HW 1 due; HW 2 assigned
- Review of HW 1

Week 03, 09/08: Resampling Methods

- ISLR Chapter 5
- Lab: Resampling Methods (ISLR 5.3)
- HW 2 due; HW 3 assigned

- Project 1 due; Project 2 assigned
- Review of HW 2

Week 04, 09/15: Linear Model Selection

- ISLR Chapter 6
- Lab: Regularization (ISLR 6.5, 6.6 & 6.7)
- HW 3 due; HW 4 assigned
- Review of HW 3

Week 05, 09/22: Moving Beyond Linearity

- ISLR Chapter 7
- Lab: Moving Beyond Linearity (ISLR 7.8)
- HW 4 due; HW 5 assigned
- Review of HW 4

Week 06, 09/29: Tree-based Methods

- ISLR Chapter 8
- Lab: Decision Trees (ISLR 8.3)
- Project 2 due; Project 3 assigned
- HW 5 due; HW 6 assigned
- Review of HW 5

Week 07, 10/06: Support Vector Machines

- ISLR Chapter 9
- Support Vector Regression (Handouts)
- Lab: Support Vector Machines (ISLR 9.6)
- HW 6 due; HW 7 assigned
- Review of HW 6

Week 08, 10/13: Fall Recess

Week 09, 10/20: Neural Networks & Deep Learning I

- The Perceptron (Handouts)
- ISLR Chapter 10
- HW 7 due;
- Review of HW 7

Week 10, 10/27: Neural Networks & Deep Learning II

- Introduction to *keras*

Week 11, 11/03: Neural Networks & Deep Learning III

- Lab: Deep Learning (ISLR 10.9)
- HW 8 assigned
- Project 3 due; Project 4 assigned

Week 12, 11/10: Unsupervised Learning

- ISLR Chapter 12
- HW 8 due; HW 9 assigned
- Review of HW 8
- Lab: Unsupervised Learning

Week 13, 11/17: Thanksgiving Holiday

Week 14, 11/24: Probability calculus, Bayesian networks: syntax and semantics

- From Propositional to Graded Beliefs (Handouts)
- Updating Beliefs (Handouts)
- Capturing Independence Graphically (Handouts)
- Parameterizing the Independence Structure (Handouts)
- HW 9 due; HW 10 assigned

Week 15, 12/01: Bias in ML

- Barocas, Solon, Moritz Hardt, and Arvind Narayanan. "Fairness in machine learning." Nips tutorial 1 (2017): 2. Available at: `https://fairmlbook.org/`. Chapters 1, 2 &4

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.

IMPORTANT: COVID-19 PROTOCOLS

Students must comply with all COVID-19 safety protocols outlined by federal, state, local, and university policies. These policies will likely evolve with the changing conditions of the COVID-19 pandemic and may include social distancing, the use of face coverings at all times, proof of vaccination, and regular COVID testing, among others. Depending on the policies outline by the above authorities, and the conditions of the class, the class might switch between meeting online and in person.