Course Number: PH240C Machine learning and biostatistics in healthcare

Course Syllabus (Fall 2022)

Table of Contents

Return to Table of Contents

Course Information

Course Meeting Dates and Times: Wednesdays 6-9 pm

Course Location: Berkeley, Berkeley Way West Building

Instructor: Jingshen Wang, PhD

Phone: (510) 643-8083

E-mail: jingshenwang@berkeley.edu

Instructor Availability: provide information as to how and when students can best contact you

GSI (if applicable): Waverly Wei, Master in Biostatistics

E-mail: linqing_wei@berkeley.edu

Course Unit Value: 4

Course Description

Machine learning (ML) algorithms are widely applied in our daily lives. The overarching goal of this course is to provide students with an overview and hands-on experiences of popular machine learning methods and biostatistical models adopted in the healthcare system and medical research. The topics of the class includes supervised learning methods (GLM, SVM, metric learning, tree-based approaches), semi-supervised learning (transduction learning, inductive learning), deep learning and neural networks, gradient descent and stochastic gradient descent, streaming data analysis, adaptive experiments, reinforcement learning and multi arm bandit algorithm. The course will also cover the applications of these methods in analyzing electronic medical record data and genetic datasets.

Prerequisites

Linear model, probability, linear algebra, and undergraduate statistics, coding experience in R, Python or any programming language

Course Learning Objectives

After successfully completing this course, you will be able to:

1. Apply popular ML methods for analyzing large-scale biomedical databases including electronic medical record data.
2. Understand and be able to interpret results obtained from ML and biostatiscal methods
3. Apply existing adaptive experimental design methods in real world situations

Methods of Instruction

Lecture and Lab sessions

Instructor Information

Jingshen Wang **Email: jingshenwang@berkeley.edu** **OH: by appointment**	

GSI: Waverly Wei **Email: linqing_wei@berkeley.edu** **OH: Monday 1-2 pm in person and zoom**
Waverly is a PhD student in Biostatistics. She has experience working with biomedical data, developing statistical methodologies under causal inference framework, and integrating machine learning methods into real world contexts. Her current research interests lie in the intersection of causal inference, reinforcement learning, and adaptive design.

Course Format

Regular lecture and Lab sessions

Return to Table of Contents

Course Schedule

Week	Dates	Topic
1	8/24	Course overview
2	8/31	Overview, supervised learning introduction, and generalized linear models
3	9/7	Support vector machines and duality theory
4	9/14	Metric learning and tree-based methods
5	9/21	Bagging and boosting
6	9/28	Overview of semi-supervised learning
7	10/5	Inductive Learning and applications in electronic medical records
8	10/12	Deep learning and neural networks
9	10/19	Gradient descent and stochastic gradient descent
10	10/26	Overview of reinforcement learning and adaptive experiments
11	11/2	Nature's experiments: Mendelian Randomization
12	11/9	Multi-armed bandit problems
13	11/16	Covariate adaptive designs and response adaptive designs
14	11/23	Class Project Presentation
15	11/30	Non-Instructional Day
16	12/7	Class Project Presentation

Course Grading

Final grades will be assigned according to the following percentages:
For example: Class Project Presentations 60%
Written assignments 30%
Class discussion/participation 10%

Course Materials

Bcourses website

To access the course website, go to bCourses at bcourses.berkeley.edu or https://phcourse.github.io/ph240c/. There you will find links to required and optional readings, the syllabus, assignment descriptions and additional course resources. Any changes will be reflected in the assignment section of the site.

Lectures

PDF versions of lecture notes will be posted on bCourses. Overview of lectures (organized by topic; some lectures may span more than one class period.

Return to Table of Contents

Course Requirements

Assignments

Assignments must be turned in for each student. An electronic copy of your assignments should be sent to your GSI by 9pm on the due date. Students can work in groups to prepare their assignments, but every student should turn in their own completed assignment.

Final Group Project

Students will identify a real data analysis project involving the causal effects of a point treatment intervention on an outcome. Groups should include up to 3 students, with one final project per group (for exceptions please contact the instructor with direct requests). Students will work with their group to implement the methods learned in class using either a provided dataset or a dataset on your own. See final project specific files on bcourses for additional description and guidance. Start thinking now if you have a data set that might be appropriate

Course Communication

As we move through the course materials, we want to hear how the course is going for you, your questions as well as how your personal and professional experiences add to our conversation. You can learn a lot from discussing the material in this course with each other and we encourage you to take advantage of the interactive components of the course to learn from each other.

Announcements

Announcements will be posted on the home page of the bCourse site. Please check regularly for updates.

Course mail

Course announcements will also be sent out through Canvas' notification system. The default is to receive announcements via the Course Mail system, so make sure to check your Course Mailbox for messages or wherever you receive notifications.

Office hours

Jingshen Wang: by appointment

Waverly Wei: Monday 1-2 pm

Policies

Due Dates

Please communicate with instructors using Canvas Course Mail if you will not be able to meet course deadlines ahead of the deadlines.

Honor Code: Individual vs. Group work

Students are expected to complete their own work. Group collaboration/discussion is permitted on all assignments. However, consulting from prior answer keys is not. Each student should write up his or her own answers to R and discussion assignments in his or her own words. The final project is collaborative; however, specific student contributions should be specified.

Course Enrollment

Participation in classroom discussions is a key component of the course.

S/U Grading

PH 240C is offered for a letter grade; if students elect to take the class S/U a Satisfactory grade (B-) requires at least 70% of total points awarded.

Correspondence

For all email correspondence please put PH 240c in the subject heading. We will try to answer your query within 48 hours during the week. If you send an email on Friday after 5:00 you will get a reply the following Monday. GSI(s) will not respond to email after 5pm the night before the assignment is due.

SPH Course Policies

Descriptions of and relevant campus links to SPH school wide course policies on Disability Support Services, Accommodation of Religions Creed, Course Evaluations, Academic Integrity can be found at: https://berkeley.box.com/s/knh3rbk9ikgvmca4ymy93msgj9bkebq5

Return to Table of Contents

Dedication to Diversity, Equity and Inclusion

Berkeley Public Health strives to create a learning environment that is inclusive of students of all backgrounds and identifies. We strive in this course to highlight how our study, work, and interactions with others can be done with empathy, a striving for understanding, and used for improvement of us as people, of our institutions and the members of our communities. We acknowledge that we will make mistakes as we are all learning together. Your perspective is important to us and the knowledge you bring to help make this an enriching learning environment for all participants. We welcome input at any time and invite constructive feedback on any suggested modifications that may help improve the course now or in the future.

Students are the experts of their own experiences. Your world lens is welcomed and encouraged; and as students, you are invited to lift up information and data that is relevant to the course material. We cannot speak on behalf of all groups, or fully understand the issues, concerns and history of students from diverse backgrounds and identities. However, we are willing to listen and learn, admit mistakes and engage in the ongoing work of intellectual humility. Racism, stubborn inequities and the continuing degradation of our environment result in injustices, which are perpetuated by silence. We commit to try to turn uncomfortable conversations into teachable moments, and invite all students to do the same even though we may not all be confident or fully-skilled in doing so. Language or comments that alienate, demean, and denigrate other students in the classroom will not be ignored, but confronted.

We also recognize that statisticians, some very well-known ones, have had a history that includes the application and advocacy of statistics with racist ends. Thus, even if our discipline would appear, at most, only tangentially related to racism and other forms of injustice, history has proved differently. We have an obligation as data scientists, to apply our knowledge with a clear eye towards the inherent biases of ourselves and others. We will study concepts of statistical bias and how rigorous application of statistical theory can result in reproducible and robust findings. In addition, we commit to ``the overarching general goal of fairness and fiercely committed to constant vigilance and scrutiny of our personal biases" (the Mismeasurement of Man; 2nd ed, 1996). In this way, we can both do good science, make meaningful contributions to a better society and help to improve the well-being of life on our fragile planet.

Return to Table of Contents