THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SYLLABUS: STAT 4620

INTRODUCTION TO STATISTICAL LEARNING

AUTUMN 2021

Course overview

Instructor

Instructor: Shili Lin
Email address: shili@stat.osu.edu
Class website: http://carmen.osu.edu
Lectures: Via CarmenZoom: Wednesdays and Fridays 1:50-2:45 pm
Office hours: Via CarmenZoom: Wednesdays 2:50-3:50 pm, Fridays 11:00-noon

Grader

Mr. Lingfei Zhao, zhao.2412@osu.edu
Tutor hours: Via Carnmen Zoom: Thursdays and Fridays 4:10-5:10 pm

Course description

This course provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimensionality reduction, and feature extraction. Specific topics include overview of predictive modeling and model evaluation, penalized regression and nonparametric regression, classification and regression trees, model selection and validation, and high-dimensional data and variable selection.
Prerequisites: STAT 3302.

Course learning outcomes

By the end of this course, students should be able to successfully:

- Recognize the types of learning problems and understand their statistical formulations.
- Understand the foundational principles of statistical learning including statistical modeling, computation, and evaluation.

- Comprehend the rationale and algorithms behind statistical learning techniques and know their relative merits and limitations.
- Evaluate and compare different learning techniques numerically in terms of generalization error.
- Use statistical learning methods for data analysis and interpret the results in the context of the data problem.

Course materials

Required

James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with Applications in R. Springer (electronic version on Carmen)

Course technology

For help with your password, university e-mail, Carmen, or any other technology issues, questions, or requests, contact the OSU IT Service Desk. Standard support hours are available at https://ocio.osu.edu/help/hours, and support for urgent issues is available 24x7.

- **Self-Service and Chat support:** http://ocio.osu.edu/selfservice
- **Phone:** 614-688-HELP (4357)
- **Email:** 8help@osu.edu
- **TDD:** 614-688-8743

Baseline technical skills necessary for online courses

- Basic computer and web-browsing skills
- Navigating Carmen

Technology skills necessary for this specific course

- CarmenZoom
- Collaborating on CarmenCanvas
- Recording a slide presentation with audio narration
- Recording, editing, and uploading video

Necessary equipment

- Computer: current Mac (OS X) or PC (Windows 10+) with high-speed internet connection
- Webcam: built-in or external webcam, fully installed
- Microphone: built-in laptop or tablet mic or external microphone

Necessary software

- This class requires you to use the free statistical software package R (The R Project for Statistical Computing; http://www.r-project.org/).
    - You can download R for Windows, Mac, and Linux, from the CRAN archive at https://cran.r-project.org.
    - An in-depth introduction to R is available at http://cran.r-project.org/doc/manuals/R-intro.pdf
    - Hands-on tutorials are available in the Swirl system, which you can learn about at http://swirlstats.com/. In particular, "R Programming: The basics of programming in R" is an appropriate first tutorial for students who have never used R.
- An easier to use interface to R is available in the free software package RStudio. This package is available for Windows, Mac, and Linux and can be downloaded from http://rstudio.org. **Note that RStudio requires R to be installed.**

Course delivery

We will meet at our regularly scheduled class time throughout the semester for most of the lectures through CarmenZoom. Recorded lectures will be posted soon after on class website. For certain topics, however, lectures may be pre-recorded and posted for asynchronous learning – learn at your own speed and at a time of your choosing – such as R tutorials and introductions to certain R packages. **Detailed instructions for asynchronous learning will be posted prior to the commencing of such activities.**

The instructor will hold weekly office hours – times as given above – via CarmenZoom.

Grading and faculty response

Homework, Projects, and Exams

Assignment or category	Percentage
Participation	5
Homework	20
Exams	45
Final Project	30
Total	100

Homework. There are six homework assignments throughout the semester. You may discuss with other students, but DO NOT simply copy any part of someone else's work or solutions from any other sources. Violations will be treated as academic misconducts. I would encourage you to talk to me if you have questions after serious attempts have been made to work on an assignment.

Exams. There are two one-hour long exams, with the tentative dates given below; any date changes will be communicated well in advance. Each exam will be proctored through webcam. The course project will serve as a cumulative evaluation of your learning in lieu of a final examination.

Exam 1	Friday, October 1	1:50-2:45 pm
Exam 2	Friday, November 12	1:50-2:45 pm

Final project. Teams of 4-5 students will be formed to work on a Final Project. Details will be introduced in class mid-way through the semester, and the project will be due near the end of the semester. Students are strongly advised to work on their project early. The teams will make recorded presentations of up to 15 minutes for peer-viewing and evaluation; they will also pitch their project in the last week of classes in a "rapid fire" round. Your overall project score will be based on the evaluations of the following components: relevant questions, appropriate analyses, communication of findings (both in the rapid fire round and the recorded project presentation), and thoughtful comments on at least two other groups' video presentations.

Logistics and policies. Homework and projects will be submitted through the class website. Typically, no late homework/project will be accepted, and no make-up exams will be given. However, if you have an emergency that prevents you from taking an exam on the date specified, please get in touch with me as soon as possible so that we can work out a solution. For the exams, you need to work independently without any forms of assistance or communication with anyone except the proctor. A basic calculator is permitted; however, using a cell phone, tablet, laptop or any other communication device for this purpose is not permitted.

Faculty feedback and response time

The following sections delineate my intended availability throughout the course.

Grading and feedback

Sample solutions to homework assignments will be posted soon after all papers are submitted. You can generally expect feedback within **7 days**, but there may be exceptions (e.g. grader has his own exam in a particular week).

E-mail

I will reply to e-mails within **24 hours on weekdays**.

Attendance, participation, and discussions

Student participation requirements

For lectures that are delivered synchronously, you are encouraged to join the live lectures and participate in discussions. If you have to miss a live lecture, you are responsible for learning the materials discussed in the recorded lectures, and you may ask questions during virtual office hours. You are also expected to participate in a series of surveys, which will collect data for one of the homework assignments.

Discussion and communication guidelines

The following are my expectations for how we should communicate as a class. Above all, please remember to be respectful and thoughtful.

- **Writing style**: While there is no need to participate in class discussions as if you were writing a research paper, you should not use text lingo.
- **Tone and civility**: We need to strive to create a supportive learning community where everyone feels safe and people can disagree amicably. Remember that sarcasm doesn't always come across online.
- **Citing your sources**: When we have academic discussions, please cite your sources to back up what you say. (For the textbook or other course materials, list at least the title and page numbers. For online sources, include a link.)
- **Backing up your work**: Consider composing your academic posts in a word processor, where you can save your work, and then copying into the Carmen discussion.

Other course policies

Health and safety

The Ohio State University Wexner Medical Center's Coronavirus Outbreak site (https://wexnermedical.osu.edu/features/coronavirus) includes the latest information about COVID-19 as well as guidance for students, faculty and staff. Guidelines and requirements for campus safety are available at the OSU Safe and Healthy website (https://safeandhealthy.osu.edu).

Student academic services

Student academic services on the OSU main campus: http://advising.osu.edu/welcome.shtml.

Student support services

Student support services offered on the OSU main campus: http://ssc.osu.edu.

Academic integrity policy

- **Reusing past work**: In general, you are prohibited in university courses from turning in work from a past class to your current class, even if you modify it. If you want to build on past research or revisit a topic you've explored in previous courses, please discuss the situation with me.
- **Falsifying research or results**: All research you will conduct in this course is intended to be a learning experience; you should never feel tempted to make your results or your library research look more successful than it was.
- **Collaboration and informal peer-review**: While study groups are encouraged, remember that copying solutions from another student or from any other sources is not permitted. If you're unsure about a particular situation, please feel free just to ask ahead of time.
- **Group projects**: This course includes a group project, which provides you the opportunity to formally collaborate with your colleagues. Please remember to work on your group project at a distance, not in person. To allow group members to take ownership of their own work, each member will be required to complete a survey (which will not be visible to the rest of their team) after project presentations to specifically delineate their individual contributions to the project.

Ohio State's academic integrity policy

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed and illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487).

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on title IX

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been

sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator, Kellie Brennan, at titleix@osu.edu

Accessibility accommodations for students with disabilities

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Accessibility of course technology

This online course requires use of Carmen (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use the following list of technologies, please request accommodations with your instructor.

- Carmen (Canvas) accessibility
- Streaming audio and video
- Synchronous course tools

Your mental health

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614- 292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273- TALK or at suicidepreventionlifeline.org

Disclaimer

This syllabus should be taken as a fairly reliable guide for the course content. However, you cannot claim any rights from it and in particular we reserve the right to change due dates, exam dates or the methods of grading and/or assessment if necessary. Any changes will be communicated to you through official course announcements.

Course schedule (tentative)

Note: Reading assignments are from the text by James et al.

Week	Dates	Topics	Assigned Readings
1	8/25, 8/27	Introduction; Linear Regression: SLR, MLR, Geometry, Loss, Weighted Least Squares	Chapters 1-3
2	9/1, 9/3	Beyond Linear Regression; Cross-Validation	Chapter 3, Sec. 5.1
3	9/8, 9/10	Classification: Logistic Regression, Linear Discriminant Analysis (LDA)	Chapter 4
4	9/15, 9/17	Bootstrap, Cross-Validation & Bootstrap Examples	Chapter 5
5	9/22, 9/24	Regularization: Ridge Regression, LASSO	Chapter 6
6	9/29	High Dimensional Data Analysis and the Curse of Dimensionality	Chapters 6
6	10/1	Midterm 1	
7	10/6, 10/8	Smoothing Splines; Local Regression	Chapter 7
8	10/13	Generalized Additive Models (GAM)	Chapters 7
8	10/15	Autumn break – no class	
9	10/20, 10/22	GAM, Regression and Classification Trees	Chapter 7, 8
10	10/27, 10/29	Trees: Boosting, Random Forests	Chapter 8
11	11/3, 11/5	Trees; Clustering: K-means	Chapter 8, 10
12	11/10	K-means	Chapter 10
12	11/12	Midterm 2	
13	11/16, 11/18	Clustering: Hierarchical	Chapter 10
14	11/24	Additional topics – time permitting	Project related
14	11/26	Thanksgiving Holiday – no class	
15	12/1	Additional topics – time permitting	Project related
15	12/3	Final Project Presentations	
16	12/8	Final Project Presentations	