THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

SYLLABUS: STAT 4620
INTRODUCTION TO STATISTICAL LEARNING
AUTUMN 2022

Course overview

Instructor

Instructor: Dr. Sebastian Kurtek
Email address: kurtek.1@stat.osu.edu
Phone number: 614-292-0463 (contact via e-mail is highly preferred)
Office hours: Wednesdays, 10AM-12PM
Office location: Cockins Hall 440B

Graduate teaching associate: Ms. Meijia Shao
GTA email address: shao.390@osu.edu
GTA office hours: TBD

Course description

STAT 4620 provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimension reduction, and feature extraction. Specific topics that will be covered 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.

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.

Prerequisites

C- or better in STAT 3302: Statistical Modeling for Discovery II

Course materials

Required Textbook

James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning with Applications in R. Springer

The textbook is available for free through SpringerLink: https://link.springer.com/book/10.1007/978-1-4614-7138-7.

I also posted an electronic copy of the book on our Carmen page.

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 this course

- Basic computer and web-browsing skills
- Navigating Carmen
- CarmenZoom

Necessary equipment

- Computer: current Mac (OS X) or PC (Windows 10+) with high-speed internet connection.

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.rproject.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

- This class will take place in-person two times per week on Wednesdays and Fridays, 1:50PM-2:45PM in Lazenby Hall 002.
- All assignments will be posted on the Carmen class website. You will be given ample time to complete the assignments. Assignment due dates will be announced in class and on the Carmen course webpage.
- I will hold weekly in-person office hours on Wednesdays, 10AM-12PM, in my office in Cockins Hall 440B.

Grading and faculty response

Grades

Assignment or category	Percentage
Participation in Discussions	**5**
Homework	**15**
Exam 1	**25**

Assignment or category	Percentage
Exam 2	**25**
Final Project	**30**
Total	**100**

All course grades will be recorded on the class website (Carmen).

Assignment information

Participation in Discussions

Discussion topics will be posted regularly on our Carmen course page. You are expected to contribute to all discussions.

Homework

I will assign homework (approximately) bi-weekly throughout the semester. You may discuss the homework with other students, but DO NOT copy any part of someone else's work or solutions from any other sources. Violations of this policy will be treated as academic misconduct. I encourage you to talk to me or the GTA (during office hours) if you have questions after serious attempts have been made to work on an assignment.

Please write clear and detailed answers to the homework problems and provide a statement interpreting the obtained results, if appropriate. You need to show your justification for or work on each homework problem. **Answers without work will not receive full credit**.

Exams

There will be two one-hour in-class exams, with tentative dates listed on the schedule at the end of the syllabus; any date changes will be communicated well in advance in class and on the Carmen webpage. The course project will serve as a cumulative evaluation of your learning in lieu of a final examination. During the exams, you will need to work independently without any form of assistance or communication with anyone. A basic calculator is permitted; however, using a cell phone, tablet, laptop or any other device for this purpose is not permitted.

Final Project

Students will be responsible for completing a team project. Each team will consist of 4-5 students. Details will be provided in class mid-way through the semester, and the project will be due near the end of the semester. The project will consist of formulating questions that can be answered with data, and performing an appropriate analysis to answer the questions. Each team will be asked to record a presentation of their project of up to 15 minutes in length for evaluation. Each team will also give a brief introduction of their project in the last week of classes in a “rapid fire” presentation session. The overall project score will be based on the

evaluation of the following components: formulation of relevant questions, appropriate analyses, and communication of findings (both in the “rapid fire” session and the recorded project presentation).

Late assignments

Homework and project assignments will be submitted via the Carmen course webpage. Generally, late assignments will not be accepted, and no make-up exams will be provided. However, if there are extenuating circumstances beyond your control, please contact the course instructor immediately.

Grading scale

93–100: A
90–92.9: A-
87–89.9: B+
83–86.9: B
80–82.9: B-
77–79.9: C+
73–76.9: C
70 –72.9: C-
67 –69.9: D+
60 –66.9: D
Below 60: E

Faculty feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course. (Remember that you can call **614-688-HELP** at any time if you have a technical problem.)

Grading and feedback

For homeworks and exams, you can generally expect feedback within **7-14 days**.

E-mail

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

Attendance and participation

Student participation requirements

Your participation is based on your in-person attendance. The following is a summary of everyone's expected participation:

- **Logging in**: **AT LEAST ONCE PER WEEK**
  Be sure you are logging in to the course in Carmen each week, including weeks with holidays. (During most weeks you will probably log in many times.) If you have a situation that might cause you to miss an entire week of class, discuss it with me *as soon as possible.*
- **In-person class meetings**: **REQUIRED**
  You are required to attend all in-person lectures and you are responsible for all material presented during these lectures. However, formal attendance will not be taken during the class.
- **Office hours**: **OPTIONAL**
  My office hours will be held in-persion in my office in Cockins Hall 440B. If you would like to discuss an assignment with me outside of my scheduled office hours, please contact me at the beginning of the week (as soon as possible).

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**: Let's maintain a supportive learning community where everyone feels safe and where 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 Cornavirus 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 from the University's COVID-19 Transition Task Force can be found on the Safe and Healthy website (https://safeandhealthy.osu.edu).

Potential disruptions to instruction

- As much as is possible, students will have access to material online if they are unable to attend class because of positive diagnosis, symptoms, or quarantine required following contact tracing.
- If the instructor is unable to be present in person because of positive diagnosis, symptoms, or quarantine following contact tracing, the course will temporarily shift to online instruction. Details will be given on the course website if this arises.

Student academic services

Student academic services offered 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

Policies

- **Homework and project assignments**: You are expected to produce original and independent work for homework and project assignments. Although students are often encouraged to work together on homework assignments, all students must submit their own written work **in their own words**. Note that allowing others to copy your work is considered academic misconduct. Academic misconduct will not be tolerated and will be dealt with procedurally in accordance with University Rule 3335-31-02. (This policy can be found at http://oaa.osu.edu/coam.html.)
- **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**: The course includes many opportunities for formal collaboration with your classmates. While study groups and peer-review of major written assignments is encouraged, remember that comparing answers on an assignment is not permitted. If you're unsure about a particular situation, please feel free just to ask ahead of time.

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; 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). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

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 (Recommended)

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. In light of the current pandemic, students seeking to request COVID-related accommodations may do so through the university's

request process, managed by Student Life Disability Services. 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; http://slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Accessibility of course technology

This course requires use of Carmen (Ohio State's learning management system). If you need additional services to use these technologies, please request accommodations with your instructor.

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 or the methods of grading and/or assessment if necessary. Any changes will be communicated to you through official course announcements.

Course schedule (tentative)

Week	Dates	Topics	Assigned Readings
1	8/24, 8/26	Introduction; Linear Regression: SLR, MLR, Geometry, Loss, Weighted Least Squares	Chapters 1-3
2	8/31, 9/2	Beyond Linear Regression; Cross-Validation	Chapter 3, Sec. 5.1
3	9/7, 9/9	Classification: Logistic Regression, Linear Discriminant Analysis (LDA)	Chapter 4
4	9/14, 9/16	Bootstrap, Cross-Validation & Bootstrap Examples	Chapter 5
5	9/21, 9/23	Regularization: Ridge Regression, LASSO	Chapter 6
6	9/28	Regularization: Dimension Reduction, PCA, PCR	Chapter 6
6	9/30	Exam 1	
7	10/5, 10/7	High Dimensional Data Analysis and the Curse of Dimensionality; Spline Regression; Smoothing Splines; Local Regression	Sec. 6.4, Chapter 7
8	10/12	Smoothing Splines; Local Regression	Chapters 7
8	10/14	Autumn Break – no class	
9	10/19, 10/21	Generalized Additive Models (GAM), Regression and Classification Trees	Chapter 7, 8
10	10/26, 10/28	Trees: Boosting, Random Forests	Chapter 8
11	11/2, 11/4	Trees; Clustering: K-means, Hierarchical	Chapter 8, 10
12	11/9	Exam 2	
12	11/11	Veterans Day – no class	
13	11/16, 11/18	Clustering: K-means, Hierarchical	Chapter 10
14	11/23, 11/25	Thanksgiving Holiday – no class	
15	12/1	Additional topics – time permitting	
15	12/3	Final Project Presentations	
16	12/7	Final Project Presentations	