THE OHIO STATE UNIVERSITY

STAT 4620
Introduction to Statistical Learning (28155)
Autumn 2026

Syllabus - Instructor Information

Oh-Ran Kwon

Professional Qualifications
Ph.D. in Statistics, University of Minnesota
M.S. in Statistics, Seoul National University
B.E. in Statistics, Sungkyunkwan University

Syllabus - Required or Recommended Readings

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/10.1007/978-1-0716-1418-1.

Syllabus - Course Schedule

Week	Topics
1	Introduction; Linear Regression: SLR, MLR, Geometry, Loss, Weighted Least Squares
2	Beyond Linear Regression; Cross-Validation
3	Classification: Logistic Regression, Linear Discriminant Analysis (LDA)
4	Bootstrap, Cross-Validation & Bootstrap Examples

Week	Topics
5	Regularization: Ridge Regression, LASSO
6	Regularization: Dimension Reduction, PCA, PCR
7	High Dimensional Data Analysis and the Curse of Dimensionality; Spline Regression; Smoothing Splines; Local Regression
8	Smoothing Splines; Local Regression
9	Generalized Additive Models (GAM), Regression and Classification Trees
10	Trees: Boosting, Random Forests
11	Trees; Clustering: K-means, Hierarchical
12-13	Clustering: K-means, Hierarchical, Project discussion
14	Thanksgiving Break — no class
15	Additional topics – time permitting, Final project presentation
16	Final project presentation

Course Details - Instructor Contact Details

Email: kwon.675@osu.edu (Please include the course title STAT 4620 in the title of your email.)

Course Details - Office Hours

Office Hours: Mondays, 10:15 AM–11:15 AM, Cockins Hall 340D, or by appointment.

Please use the following calendar to sign up for Monday office hours: https://calendly.com/ohrankwon/office-hours. For appointments outside the scheduled office hours, please contact me by email, preferably at the beginning of the week.

Course Details - Additional Instructional Staff

Teaching Assistant: Chenze Li (Li.13074@buckeyemail.osu.edu)

Course Details - Class Meeting Details

Class Meeting Details: 11:30 AM-12:25 PM - Pomerene Hall 150 - WF

Course Details - Course Description

The course provides an introduction to the principles of statistical learning and standard learning techniques for regression, classification, clustering, dimensionality reduction, and feature extraction. Prereq: C- or better in 3302, or permission of instructor.
Prerequisite(s) / Corequisites: Prereq: C- or better in 3302, or permission of instructor.

Course Details - More About This Course

Overview: STAT 4620 introduces the core concepts of statistical learning and common techniques used in regression, classification, clustering, dimension reduction, and feature extraction. Topics covered include an 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 Details - Course Learning Outcomes

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

1. Recognize the types of learning problems and understand their statistical formulations;
2. Understand the foundational principles of statistical learning, including statistical modeling, computation, and evaluation;
3. Comprehend the rationale and algorithms behind statistical learning techniques and know their relative merits and limitations;
4. Evaluate and compare different learning techniques numerically in terms

of generalization error;
5. Use statistical learning methods for data analysis and interpret the results in the context of the data problem.

Course Details - Detailed Course Schedule

Course Schedule (Tentative)

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

Week	Date	Topics	Reading
13	11/18, 11/20	Clustering: K-means, Hierarchical, Project discussion	Chapter 10
14	11/25, 11/27	Thanksgiving Break — no class	
15	12/2	Additional topics – time permitting	
15	12/4	**Final project presentation**	
16	12/9	**Final project presentation**	

Dates for exams:

- **Exam 1: Friday, October 2, 2026, during class time.**
- **Exam 2: Friday, November 13, 2026, during class time.**
- **Final project presentations: December 4 and 9, 2026, during class time.**

Course Details - Grading Policy

Component	Percentage
Homework	20%
Exam 1	25%
Exam 2	25%
Final project	30%
Total	100%

Homework: I will assign homework (approximately) bi-weekly throughout the semester. Homework assignments must be **combined into a single PDF file and submitted through Carmen by the due date. All homework assignments will be graded based on completeness rather than correctness of the solutions.** Solving homework problems provides important practice of the material that is covered in this course. 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 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.

Please note late submission will be penalized at a rate of 5% of the points earned on the assignment per day late. The late penalty may be waived if prior approval has been obtained. If you are unable to complete an assignment on time and need extra time, please get in touch with me as soon as possible so we can discuss your situation.

Exams: There will be two 55-minute in-class exams. 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. **Make-up exams will generally not be provided. If there are extenuating circumstances beyond your control, please contact the course instructor immediately. Appropriate documentation is required.**

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 submit a final report and 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 project report).

Course Details - Grading Scale

Letter Grade	Percentage Range
A	100% to 93%

Letter Grade	Percentage Range
A-	< 93% to 90%
B+	< 90% to 87%
B	< 87% to 83%
B-	< 83% to 80%
C+	< 80% to 77%
C	< 77% to 73%
C-	< 73% to 70%
D+	< 70% to 67%
D	< 67% to 60%
E	< 60% to 0%

Course Details - Other

Course Delivery

The course will be primarily delivered in person. Some lectures may be presented online, either live over Zoom during the scheduled lecture time or as recorded lectures posted on the class website in Carmen.

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.

Course Technology

For help with your password, university email, Carmen, or any other technology issues, questions, or requests, contact the Ohio State IT Service Desk. Standard support hours are available at it.osu.edu/help, and support for urgent issues is available 24/7.

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

Technology skills needed for this course

- Basic computer and web-browsing skills
- Navigating Carmen (go.osu.edu/canvasstudent)
- CarmenZoom virtual meetings (go.osu.edu/zoom-meetings)

Required Equipment

- Computer: current Mac (MacOs) or PC (Windows 10) with high-speed internet connection
- Other: a mobile device (smartphone or tablet) to use for BuckeyePass authentication

Required 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 https://posit.co. **Note that RStudio requires R to be installed.**

Course Details - Instructor Policies

Faculty feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course.

- **Grading and feedback:** You can generally expect graded homework assignments back within 14 days.
- **E-mail:** Please include the course title STAT 4620 in the subject line of your email. I will reply to emails related to course logistics within 72 hours on school days.

Course Details - Course Policies

Student Participation Requirements

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. 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 lectures:** REGULAR ATTENDANCE EXPECTED
  Students are expected to attend in-person lectures regularly and are responsible for all material presented during these lectures. Students are also expected to participate, discuss, and answer questions in lectures.
- **Office hours:** OPTIONAL
  All office hours are optional and will be held in person in my office.

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.

- Tone and civility: Let us maintain a supportive learning community where everyone feels safe and where people can disagree amicably.
- Citing your sources: When we have academic discussions, please cite your sources to back up what you say.

Academic Integrity 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. (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 have 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 are unsure about a particular situation, please feel free just to ask ahead of time.

Disclaimer

This syllabus should be taken as a fairly reliable guide for the course content. However, due dates, course topics, or methods of grading and assessment may be changed if necessary. Any changes will be communicated through official course announcements.

Course Details - AI Use Policy

GenAI can fabricate information (for example, creating citations to sources that do not exist) or omit key information. You are ultimately responsible for the

accuracy and integrity of your submitted work. Unless explicitly permitted below, representing AI-generated content as your own work is a violation of the Code of Student Conduct. Please check with your instructor if you have questions.

The following policy applies to all course assignments unless otherwise stated. **Use of GenAI is prohibited on exams.**

- **Permitted:** You may use GenAI as a virtual tutor to explain concepts, quiz you on course topics, or summarize content you have personally authored (e.g., your own class notes). You may also use GenAI to generate topic ideas, seek examples, brainstorm ideas, or assist with understanding coding and debugging.
- **Prohibited:** The materials in this course may be subject to copyright protection. Do not input or allow AI to access external or instructor-created course materials (e.g., assigned readings, slides) without explicit instructor permission. You must personally author all work that you submit; AI-generated answers, solutions, text, or code may not be submitted as your own work.
- **Requirement: If you use GenAI for an assignment, please include a brief statement with your submission specifying the name of the tool you used (e.g., Copilot, ChatGPT, Grammarly) and how you used it.** You are responsible for the accuracy and appropriateness of any information obtained from GenAI.

Unauthorized use of GenAI will be handled in accordance with Ohio State's academic integrity policies.

If I suspect that you have used GenAI inappropriately on an assignment for this course, I will ask you to communicate with me to explain your process for completing the assignment in question.

Syllabus - University Policies

Office of Undergraduate Education - Standard Syllabus Statements

- Academic Misconduct

- Religious Accommodations
- Accommodations for Disability and Illness
- Intellectual Diversity
- Grievances and Solving Problems
- Creating an Environment Free from Harassment, Discrimination, and Sexual Misconduct
- Copyrights
- Content Warnings
- Mental Health Support
- Resources for Military-Connected Students
- Campus-wide Class Cancellations