Syllabus

STAT 6500 Statistical Machine Learning

Spring 2026

3 Credit Hours

Course overview

Lectures

MWF 3:00pm-3:55pm in Cockins Hall 232

Instructor

- Name: Yoonkyung Lee
- Office: Cockins Hall 440H
- Office Hours: M 4:10-5:05pm, R 2:00-2:55pm, or by appointment
- Email: lee.2272@osu.edu
- Phone: 614-292-9495

Grader

- Name: Kiljae Lee
- Office: Cockins Hall 305E
- Office Hours: by appointment only
- Email: lee.10428@buckeyemail.osu.edu

Course webpage

The course has a web page on Carmen (https://carmen.osu.edu/). You will find the class schedule, course announcements, homework assignments, and other information about the class on the web page. Please check it out on a regular basis.

Course description

This course explores the methodology and algorithms behind modern supervised and unsupervised learning techniques to explore relationships between variables in large, complex datasets. Topics include linear and logistic regression, classification, clustering, resampling methods, model selection and regularization, and non-linear regression. Students will also gain exposure to popular statistical machine learning algorithms implemented in R. A focus will be on understanding the formulation of statistical models and their implementation, and the practical application of learning methods to real-world datasets.

Course expected learning outcomes

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

1. Students will understand the statistical learning framework, including core concepts such as loss, learning, and generalization; they will be able to judge when the framework is applicable and be able to formulate problems within this framework.

2. Students will recognize the role of statistical models that are appropriate for a variety of statistical learning problems; they will understand the assumptions, formulation, and evaluation of these models.

3. Students will understand the rationale and algorithms behind statistical learning methods, and they will know the merits and limitations of these methods.

4. Students will be able to quantitatively evaluate and compare different statistical learning methods.

5. Students will be able to apply statistical learning methods for data analysis and be able to interpret the results in the context of the application.

Prerequisites

STAT 6450 (Applied Regression Analysis) or permission of instructor. Familiarity with calculus, linear algebra and linear regression analysis will be assumed. Basic proficiency in a programming language, such as R is required.

Computing

One of the goals of the course is to train students in the computing aspects of statistical machine learning and develop the skills to implement machine learning algorithms. Many homework assignments will have a computing and programming component. There will be example codes provided, primarily written in the language R. Students are welcome to use Python and other languages.

Attendance Policy

You are expected to attend all lectures.

Textbooks

Required

James, Witten, Hastie, Tibshirani: *An Introduction to Statistical Learning with Applications in R (Python)*, 2nd edition. (Freely downloadable PDF available at https://statlearning.com/)

Recommended

Murphy: *Machine Learning: A Probabilistic Perspective* (An electronic version is available for online reading through the OSU library website)

Grading

How your grade is calculated

Category	Percentage
Homework Assignments	30%
Midterm (take home)	30%
Group Project	30%
Participation	10%
Total	**100%**

Grading Scale

While the standard grading scale generally applies, final grades may be curved upwards.

Homework assignments

Homework will be assigned regularly (about every two weeks) throughout the semester using Carmen. Assignments will consist of a mix of technical questions to assess students' understanding of the statistical models, and questions asking students to perform analyses of datasets. The grade for the analysis portion of each assignment will be based on both the accuracy and appropriateness of the analysis, as well as the clarity of the description of the analysis and results. The assigned problems and solutions will be posted on the course website.

Late assignments

Late assignments will be accepted for 24 hours after the original due date with a 1% deduction per hour. After this, no late assignments will be accepted.

Midterm

A take-home midterm exam will be given. The midterm will be completed by each student individually.

Group project

Students will also complete projects in groups consisting of 3 to 4 members (depending on the enrollment size). The project will consist of selecting a data set (by week 4), performing an exploratory data analysis (EDA, by week 8), making a 5-page proposal (by week 12), presenting (in week 15), and submitting a 10-page final report (by May 4). The proposal should contain a detailed problem statement that includes questions of interest, and a description of what methods will be used and how they will be used to answer questions of interest or solve the problem. More details will be given later.

Participation

You are expected to attend all lectures. In addition to regular class participation, there will be several activities that require your participation to build connections with other students and formulate potential projects (e.g., posting introduction video, proposing datasets for project). These activities will be announced in class and on Carmen.

Academic policies

The link below provides academic policies and statements from the Office of Undergraduate Education:
https://ugeducation.osu.edu/academics/syllabus-policies-statements/standard-syllabus-statements

Academic integrity policy

Although students are encouraged to work together on assignments, each student is expected to write and submit individual solutions to homework problems. The midterm is to be completed on your own without any

external help or communication. Academic misconduct will not be tolerated and will be dealt with procedurally in accordance with university 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 3335-5-48.7(B)). For additional information, see the Code of Student Conduct: http://studentlife.osu.edu/csc/.

Accommodations for disability or illness

The university strives to maintain a healthy and accessible environment to support student learning in and out of the classroom. If students anticipate or experience academic barriers based on a disability (including mental health and medical conditions, whether chronic or temporary), they should let their instructor know immediately so that they can privately discuss options. Students do not need to disclose specific information about a disability to faculty. To establish reasonable accommodations, students may be asked to register with Student Life Disability Services. After registration, students should make arrangements with their instructors as soon as possible to discuss your accommodations so that accommodations may be implemented in a timely fashion.

If students are ill and need to miss class, including if they are staying home and away from others while experiencing symptoms of viral infection or fever, they should let their instructor know immediately. In cases where illness interacts with an underlying medical condition, please consult with Student Life Disability Services to request reasonable accommodations.

You can connect with them at slds@osu.edu; 614-292-3307; or slds.osu.edu.

Copyright for instructional materials

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.

Course schedule (tentative)

Homework due dates and midterm schedule in the table are tentative. Please refer to in-class announcements (also on Carmen) for official dates.

Week	Date	Topics/Readings	Assignments Due
1	1/12-1/16	Intro to Statistical Learning	
2	1/19-1/23	1/19 (M): Martin Luther King Jr Day Review of Linear Regression, Classification: Logistic regression, Gaussian LDA	Project: Dataset Proposal
3	1/26-1/30	Quadratic DA, Comparison of Methods, Evaluation Criteria	Homework 1
4	2/2-2/6	Application*, Resampling methods: Cross-validation, Bootstrap	Project: Dataset Selection

Week	Date	Topics/Readings	Assignments Due
5	2/9-2/13	Linear Model Selection and Regularization: Subset Selection, Shrinkage Methods, Dimension Reduction Method	Homework 2
6	2/16-2/20	Linear Model Selection and Regularization, Basis Expansion Approach	
7	2/23-2/27	Splines, Smoothing Splines	Homework 3
8	3/2-3/6	Local Regression, Generalized Additive Models, Application*	Project: EDA
9	3/9-3/13	Support Vector Machines and Maximal Margin Classifier	Homework 4
	3/16-3/20	Spring Break	
10	3/23-3/27	Kernels for Nonlinear SVM, Application*	Homework 5 Midterm begins
11	3/30-4/3	Tree-based method: Classification and Regression Trees	Midterm due
12	4/6-4/10	Bagging, Random Forest, Boosting	Project: Proposal

Week	Date	Topics/Readings	Assignments Due
13	4/13-4/17	Application*, PCA, Matrix Completion	
14	4/20-4/24	Clustering Methods: k-Means, Hierarchical Clustering, Neural Networks	
15	4/27 (M) 5/1 (F)	Neural Networks Project presentation video due	Project: Presentation video
16	5/4 (M)	Project report due	Project: Report

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.