

**SC324 – Statistical Learning in Data Science
Spring 2025****Professor:**

Jerzy Wieczorek

Lovejoy 110

207-859-5841

jawieczo@colby.edu**Class Meetings:**

MWF: 9:00-9:50am

Lovejoy 205

Office Hours:

Posted on Moodle page.

Course Description:

Statistical methods used in data science allow computers to make inferences and predictions about target variables. This course will provide students exposure to the common statistical methods and models used in this setting. Although the emphasis is on applications, the statistical and mathematical foundations for these data science techniques will also be covered. Topics will include linear modeling and classification techniques, cross validation, bootstrapping, non-linear modeling, tree-based methods, and data reduction strategies. Unsupervised learning techniques will also be covered as time allows.

Prerequisite: Statistics 212.*Corequisite:* Mathematics 253 (may be taken concurrently).**Learning Objectives:** Students will learn to...

1. Understand broader contexts for the use of statistical learning techniques in data science applications, and how these techniques fit into the intersection among statistics, mathematics, and computer science.
2. Explain the importance of, and differences between, prediction and inference.
3. Apply cross-validation methods for model tuning and model selection, and bootstrap-based methods for quantifying precision and uncertainty.
4. Apply linear modeling and classification techniques.
5. Apply non-linear modeling techniques as an extension of linear models.
6. Use tree-based and SVM-based methods for linear and non-linear classification.
7. Explore data reduction techniques such as clustering and principal components.
8. Take appropriate considerations into account when choosing among available data science techniques, particularly: the bias-variance tradeoff, computational tradeoffs, the curse of dimensionality, and sparsity and regularization.

Assessment:

The final course grade will be determined by the following components:

Participation	5%
Labs	5%
Homework	20%
Exam 1	30%
Exam 2	30%
Project	10%

Participation (5%) will be assessed through class attendance.

Labs (5%) will be assigned from the book to get everyone up to speed using R. Labs should be completed in RStudio using RMarkdown. Both your .Rmd file and your knitted output should be uploaded through Moodle.

Homework (20%) will be assigned approximately weekly or bi-weekly with due dates posted on Moodle. This will provide an opportunity for frequent assessment of understanding of the material presented in the reading and in class. You may work together on homework, but you are expected to write up answers in your own words. ***Late homework will reduce the student's grade substantially (by 25% of the maximum score for each day late), and very late homeworks may not be accepted, in order to let the graders return everyone's work quickly and give you timely feedback. The professor will not grant extensions except for documented illness or other extreme circumstances.***

Two exams (30% each) will be held, each consisting of up to two in-class portions. The dates of these exams will be posted on Moodle. The exams are closed-book and closed-notes unless otherwise specified.

There will be a **project** (10%) consisting of using a data set of your choosing to make inferences and predictions. The final written project will be due during our final exam period.

Required Text:

James, Witten, Hastie, and Tibshirani (2021). *An Introduction to Statistical Learning*, 2nd ed. Springer. ISBN: 978-1-0716-1417-4.

Be sure you have the **second edition, with R (not Python)**. It is available at our bookstore. Alternately, you can download a free PDF of the textbook through the Colby libraries:

<https://link.springer.com/book/10.1007/978-1-0716-1418-1>

You can also purchase a cheaper black-and-white copy of the book there by clicking on "MyCopy Softcover" which is currently around \$39.99.

Statistical Software and Materials:

We will use the free statistical software package R and RStudio. Although there is a Python edition of our textbook, the instructor's materials (including most assignments) will all be in R. Our course will have its own dedicated RStudio server at <https://sc324.colby.edu/>. To access this server from off campus, you will need to use Colby's Virtual Private Network (VPN): see [What is the VPN?](#) and [Setting up VPN](#).

Course Website:

A course website will be maintained for this course. Data sets, homework assignments, and relevant handouts and resources will be posted on the Moodle website for the course, available at <https://moodle.colby.edu>.

Masks:

According to guidance from the Dean of Faculty, “Regardless of campus-wide policies, faculty and teaching staff may require masks in classrooms and other teaching spaces. More generally, we appreciate everyone respecting the needs of others and wearing masks as requested, and we ask that you have a mask with you when you are on campus.”

Depending on conditions, I may expect the SC324 classroom to mask in order to protect those of us with compromised immune systems, those of us with high-risk family members, and those of us who simply want to avoid transmitting the virus to others or being re/infected themselves.

Lectures:

Use common courtesy: arrive on time; do not leave early; no cell-phone use allowed; do not be disruptive in class; participate in class when the instructor asks questions; etc. The use of laptops/tablets/etc. is allowed only for course-related purposes; students are encouraged to follow along and run code in class.

Deadline extensions and missed classes:

Colby College supports the religious practices of students, faculty, and staff. Students are expected to notify their instructors of their intent to fulfill the obligations of their religious tradition well in advance of these days. For this class I ask that you notify me by email at least 14 days in advance of the date in question.

Colby rules state that student athletes are permitted to miss a class for a contest but only at the discretion of the professor of that class. Students should communicate with the professor directly and as soon as possible (ideally in the first week of classes) about any conflicts between their athletic schedule and our class schedule.

In general, extensions will not be granted for students because they are behind on work, had a busy week, etc. Extensions for **reasonable academic purposes** (e.g. job interview) or **extreme circumstances** (e.g. hospitalization) will generally be granted. If you believe you have a reasonable request for an extension, please request this at least 48 hours before an assignment is due, along with the reason for requesting an extension. At the top of the assignment, please clearly write that you received an extension on the assignment.

Academic accommodations:

I am available to discuss academic accommodations that any student with a documented disability may require. Please note that you'll need to provide a letter from the Dean of Studies Office documenting your approved accommodations. Please meet with me within two weeks of the start of the semester to make a request for accommodations so that we can work together with the College to make the appropriate arrangements for you. Colby's Office for Student Access and Disability Services is the primary contact for accommodations and any questions related to educational testing and documentation:

<https://life.colby.edu/get-support/access-disability-services/>

Sexual misconduct/Title IX statement:

Colby College prohibits and will not tolerate sexual misconduct or gender-based discrimination of any kind. Colby is legally obligated to investigate sexual misconduct (including, but not limited to sexual assault and sexual harassment).

If you wish to speak about an incident of sexual misconduct, please contact Colby Counseling Services (207-859-4490; confidential) or Colby's Title IX Coordinator, Emily Schusterbauer (207-859-4266; non-confidential).

Students should be aware that faculty members are considered responsible employees; as such, if you disclose an incident of sexual misconduct to a faculty member, they have an obligation to report it to Colby's Title IX Coordinator. "Disclosure" may include communication in-person, via email/phone/text, or through class assignments.

To learn more about sexual misconduct or report an incident, visit

<https://life.colby.edu/your-safety/sexual-violence-title-ix/>

Academic honesty and consequences for academic dishonesty:

Always ask if you are unsure whether your actions comply with the assignment instructions.

Always acknowledge any help received on assignments: list the names of the people you worked with, and cite any external sources you used. You are encouraged to discuss assignments with your classmates, but the work you submit must be your own.

Honesty, integrity, and personal responsibility are cornerstones of a Colby education and provide the foundation for scholarly inquiry, intellectual discourse, and an open and welcoming campus community. These values are articulated in the Colby Affirmation and are central to this course. Students are expected to demonstrate academic honesty in all aspects of this course.

Academic dishonesty includes, but is not limited to: plagiarism (including quoting sources without “ ” around the borrowed words and a citation); presenting another's work as one's own; buying or attempting to buy papers or projects for a course; fabricating information or citations; knowingly assisting others in acts of academic dishonesty; violating clearly stated rules for taking an exam or completing homework; misrepresentations to faculty within the context of a course; and submitting the same work, including an essay that you wrote, in more than one course without the permission of instructors.

Academic dishonesty is a serious offense against the college. Sanctions for academic dishonesty are assigned by an academic review board and may include failure on the assignment, failure in the course, or suspension or expulsion from the College.

For more on recognizing and avoiding plagiarism, see:

<https://libguides.colby.edu/avoidingplagiarism>

For resources and information on academic integrity, see:

<https://www.colby.edu/academics/academic-integrity/>

A special note on AI and ChatGPT:

Under no circumstances should you be using any type of AI output in your submissions for assignments or exams. This includes, but is not necessarily limited to, using any kind of online AI chat (e.g. ChatGPT). You may find AI useful as a supplementary learning tool (such as searching for code syntax, or rephrasing the explanation of a statistical concept to help you understand it yourself), but you may not use AI to produce answers to the assignments themselves. All writing that is conducted in this class must be original and be your own. If you have any questions about this, ask me.

The Colby Affirmation:

Colby College is a community dedicated to learning and committed to the growth and well-being of all its members. As a community devoted to intellectual growth, we value academic integrity. We agree to take ownership of our academic work, to submit only work that is our own, to fully acknowledge the research and ideas of others in our work, and to abide by the instructions and regulations governing academic work established by the faculty.

As a community built on respect for ourselves, each other, and our physical environment, we recognize the diversity of people who have gathered here and that genuine inclusivity requires active, honest, and compassionate engagement with one another. We agree to respect each other, to honor community expectations, and to comply with College policies.

As a member of this community, I pledge to hold myself and others accountable to these values wherever I may find myself.

<https://www.colby.edu/academics/academic-integrity/the-colby-affirmation/>