

# CSC 171 - Applied Machine Learning in R - Summer 2022

## Course Information

| Course Location | Meeting Days         | Time         |
|-----------------|----------------------|--------------|
| TBD             | Tuesdays & Thursdays | 9:30-11:30am |

## Instructor Information

| Instructor Information   | Office Location  | Hours |
|--|--|-------|
| Dr. Jason S. Byers<br><a href="mailto:jabyers@davidson.edu">(jabyers@davidson.edu)</a><br>(704) 894-2760 | Main Office<br>Chambers 2256<br><br>Data CATS, drop in hours<br>Hurt Hub | TBD   |

## Syllabus

### Course Home

Everything you need for this class (announcements, resources, assignments and other activities) will be posted on the course website. Please plan to check the page regularly.

### Course Meeting Link

TBD

### Course Format

This is an online course with both synchronous and asynchronous components. Here's what that means in practice: each week, you will be assigned a lab to work through. Each lab will typically comprise a set of readings/videos and a coding exercise to be completed. You will then meet with me via Zoom once a week to discuss the readings/videos and the coding exercises on Thursday 9:30 am - 11:30 am. Follow-up meetings may be scheduled as necessary on Tuesdays 9:30 am - 11:30 am, EST.

### Course Description

Machine learning is the subfield of Artificial Intelligence that is concerned with the problem of designing algorithms and systems that improve their performance in a certain task with accumulated experience. While the ability to learn is clearly a key trait that any system attempting to behave "intelligently" must possess, machine learning techniques have increasingly become central to many software systems. For example, learning algorithms are a fundamental component of state-of-the-art systems for filtering spam, detecting fraud, recommending products to purchase, and understanding visual and textual content. This course will introduce students to some of the fundamental algorithms in this field, the theory that underpins these approaches, and the practicalities of applying these ideas to novel, real-world problems. Topics that will be covered include techniques for regression (linear, logistical, polynomial), classification algorithms (*k*-nearest neighbors, decision trees, support vector machines), the bias-variance decomposition, ensemble methods (bagging, boosting), and dimensionality reduction techniques.

### Learning Outcomes

Together, we will strive for your individual and collective success in achieving the learning outcomes of this course. At the conclusion of this course, students will be able to:

- Describe the regression problem, understand various approaches to solving regression problems (linear, polynomial, ridge) and apply them to real-world datasets,
- Describe the bias-variance trade-off and articulate its implications for machine learning practitioners and data scientists
- Describe the classification problem, understand various approaches to solving classification problems (instance-based algorithms, decision trees, support vector machines, neural networks) and apply them to real-world datasets
- Appreciate the importance of ensemble methods like bagging and boosting and their relationship to the bias-variance decomposition
- Understand the "curse of dimensionality" and appreciate the usefulness of dimensionality reduction techniques like feature selection and principal components analysis

### Prerequisites

MAT 105 (or an equivalent) or by permission of the instructor

### Course Materials

To maximize access to this class, we will use freely available textbooks, videos, and other resources, with a focus on the following:

- Primary text (ISL): James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. *An Introduction to Statistical Learning with Applications in R*. Second Edition. New York: Springer. This book is freely available online. It is also available in paperback, if you prefer a hard copy. Warning: some content and the numbering system differs between print and online versions; I will exclusively refer to the free online version.
- Supplementary text (HOML): Boehmke, Bradley, and Brandon Greenwell. 2020. *Hands-On Machine Learning with R*. New York: Chapman and Hall/CRC. This book is freely available online. It is also available in paperback, if you prefer a hard copy. Warning: some content and the numbering system differs between print and online versions; I will exclusively refer to the free online version.

### Software

You will use two freely available programs, R and RStudio, in order to complete the assignments for this course. R and RStudio are installed on all Davidson campus computers. They are also freely available to install on your own computer.

### Access and Accommodation

The college welcomes requests for accommodations related to disability and will grant those that are determined to be reasonable and maintain the integrity of a program or curriculum. To make such a request or to begin a conversation about a possible request, please contact the Office of Academic Access and Disability Resources, which is located in the Center for Teaching and Learning in the E.H. Little Library: Beth Bleil, Director, [blebel@davidson.edu](mailto:blebel@davidson.edu), 704-894-2129; or Alysen Beatty, Assistant Director, [albeaty@davidson.edu](mailto:albeaty@davidson.edu), 704-894-2939. It is best to submit accommodation requests within the drop/add period; however, requests can be made at any time in the semester. Please keep in mind that accommodations are not retroactive.

### Course Organization

Modes of learning in this class (whether assessed directly or indirectly) require a range of skills and abilities. Every student's success is important to me, and I am happy to work with you to develop strategies for success in this class. For Summer 2021, we will be meeting remotely, to allow everyone to participate fully in the collaborative environment that is necessary to maximize your learning.

- **In-Class Activities.** Each class day will involve a significant amount of discussion of the readings and topics for that week. Additionally, the class meeting will be a time for students to demonstrate their applied knowledge of the readings and topics for each week. In order for these activities to be effective, you must do the assigned readings and videos before you come to class, and be prepared to ask (and answer) questions before diving into a discussion.
- **Labs.** Weekly assignments (due approximately every Thursday at 9:30 am EST, for a total of 10 labs) will provide you with regular practice applying machine learning techniques in R. These assignments will build on the material presented in class, and require you to apply the concepts in new ways.
- **Final Project.** The goal of the final project is for you to apply the machine learning techniques and skills learned in this course to real data.

### Attendance Policy

Missing class will adversely affect your grade in many ways. In addition, the college attendance policy will be enforced: missing more than 25% of class meetings makes you eligible for a failing grade. Please look carefully at the syllabus during the first week of class. Should there be a conflict between any class session or assignment due date and a religious holiday or observance, athletic contest, or another academic or personal commitment please let me know well in advance. Religious observance warrants a legitimately excused absence. If you must miss class for any reason, excused or otherwise, you are responsible for getting notes from a classmate and turning in all work on time. Each student will be granted 2 unexcused absences.

### Getting Help

It is normal and expected that all students will need help outside of class with the material in this course. Because a language like R is only learned with practice, an important source of help is additional exercises, in the required textbook or optional online resources provided on the course web page. The following additional resources are also available.

- **Office Hours.** I welcome you to visit me during the hours listed at the beginning of this document. It is a good practice to make an appointment with me even if outside of the listed hours of availability.
- **Reusing/Sharing Code.** Many of the datasets we will discuss and analyze are publicly available, so they may have been extensively discussed and analyzed. Unless explicitly instructed otherwise, you may use available code and resources for course activities (e.g., Github repos, StackOverflow answers) but you must cite the source of the code/resource within your program files and/or document. Recycled code that is discovered that is not properly cited may be considered as plagiarism. When working in groups on class assignments you are welcome to discuss problems together and ask for general advice, but you may not share or use code from another group.
- **Honor Code.** Please adhere to the Davidson College Honor Pledge.

### Grading

| Category      | Points    |
|---------------|-----------|
| Attendance    | 10 Points |
| Participation | 10 Points |
| Labs          | 60 Points |
| Final Project | 20 Points |

## Schedule

A tentative class schedule of topics, readings and due dates is available below. Minor adjustments will be made as needed, on the course web page. Please double check the web page before doing each reading assignment.

### Week 1

#### Topics

- Introduction
- What is statistical learning?

#### Date    Readings

6/2    [ISL Chapter 1 & ISL Chapter 2](#)

#### Assignments

- Lab 1 Assigned

### Week 2

#### Topics

- Linear Regression

#### Date    Readings

6/9    [ISL Chapter 3](#)

#### Assignments

- Lab 2 Assigned
- Lab 1 DUE

### Week 3

#### Topics

- Classification

#### Date    Readings

6/16    [ISL Chapter 4](#)

#### Assignments

- Lab 3 Assigned
- Lab 2 DUE

### Week 4

#### Topics

- Resampling Methods

#### Date    Readings

6/23    [ISL Chapter 5](#)

#### Assignments

- Lab 4 Assigned
- Lab 3 DUE

### Week 5

#### Topics

- Linear Model Selection and Regularization

#### Date    Readings

6/30    [ISL Chapter 6](#)

#### Assignments

- Lab 5 Assigned
- Lab 4 DUE

### Week 6

#### Topics

- Moving Beyond Linearity

#### Date    Readings

7/7    [ISL Chapter 7](#)

#### Assignments

- Lab 6 Assigned
- Lab 5 DUE

### Week 7

#### Topics

- Tree-Based Methods

#### Date    Readings

7/14    [ISL Chapter 8](#)

#### Assignments

- Lab 7 Assigned
- Lab 6 DUE

### Week 8

#### Topics

- Support Vector Machines

#### Date    Readings

7/21    [ISL Chapter 9](#)

#### Assignments

- Lab 8 Assigned
- Lab 7 DUE

### Week 9

#### Topics

- Deep Learning

#### Date    Readings

7/28    [ISL Chapter 10](#)

#### Assignments

- Lab 9 Assigned
- Lab 8 DUE

### Week 10

#### Topics

- Unsupervised Learning

#### Date    Readings

8/4    [ISL Chapter 12](#)

#### Assignments

- Lab 10 Assigned
- Lab 9 DUE

### Week 11

#### Topics

- Final Project

#### Date    Readings

8/11    Final Project

#### Assignments

- Lab 10 DUE