STA 4241

Fall 2024

Statistical Learning in R

Instructor: Joseph Antonelli
E-mail: jantonelli@ufl.edu
Office hours: Tuesdays from 10:45am to 11:45am, and Wednesdays from 2:00pm to 3:00pm. Both are in Griffin-Floyd Hall 206.

Teaching Assistant: Zeyu Yuwen
E-mail: zeyu.yuwen@ufl.edu
Office hours: TBD

Course Website: e-Learning

Course Prerequisites: STA 4322 & STA 4210 & MAS 4115 (Or linear algebra equivalent)

Course lectures: Tuesdays from 11:45am to 1:40pm and Thursdays from 12:50pm to 1:40pm. Both lectures are held in Griffin-Floyd Hall 100.

Course Notes/Material: Notes for the week will be posted at the beginning of each week on the course website. These should contain nearly all of the material that we cover in class, however, I will present some additional material in the lectures that is not posted on the course website.

Software: We will be using the R software language throughout. R is free and should be easy to download on your personal computer. I also highly recommend running R through RStudio, though it is not a requirement. If you have any problems downloading R or RStudio, feel free to talk to myself or the TA. If you do not have access to a computer, please reach out to me via email.

Required Text: James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) An Introduction to Statistical learning with Applications in R. Springer.

ISBN-13: 978-1461471370

Course Description: Overview of the field of statistical learning. Topics include linear regression, classification, resampling methods, shrinkage approaches, nonlinear regression, tree-based methods, support vector machines, and neural networks. We will cover many aspects of these approaches including conceptual, theoretical, and applied aspects. Approaches will be illustrated and implemented in R.

Course Objectives: The goal of this course is to teach the theoretical underpinnings of a number of advanced and commonly used statistical learning techniques. We will review classical statistical techniques such as linear and logistic regression before covering more advanced statistical techniques such as classification, regularization, nonlinear regression, and other machine learning approaches. The implementation of all approaches in the R statistical software will be taught throughout. By the end of the course, students should be familiar with a wide range of statistical methodologies that are widely used in practice, and should be able to apply these approaches to data sets.

Homework

There will be a homework assignment roughly every two weeks and it will be due via Canvas submission.

Exams

You will have one take-home exam that is to be assigned roughly in the middle of the semester.

Project: Students will be expected to complete a written project at the end of the semester and present their findings to the class. The grade for the final project will consist of two main components: a written report and an oral presentation. The written report will state the objectives of the study, describe data collection, describe the statistical model used, explain any assumptions required by the analysis, and provide conclusions for the main study questions. The oral presentation will be a 10-15 minute presentation that should cover the key components of the written report. Presentations should clearly state the objectives of the project, while using visualizations to illustrate the main results and findings of the project.

Grade Distribution

Homework	20%
Midterm	40%
Project	40%

Letter Grade Assignment: Grades will be assigned as follows: 90-100, A; 87-89.9, A-; 84-86.9, B+; 80-83.9, B; 77- 79.9, B-; 74-76.9, C+; 70-73.9, C; 67-69.9, C-; 64-66.9, D+; 60-63.9, D; 55-59.9, D-; 0- 55, F

The numeric scores will be rounded to the nearest tenth.

Make up Policy: Requirements for class attendance and make-up exams, assignments, and other work in this course as well as policies regarding absences, religious holidays, illness and student athletes are consistent with UF Attendance Policies (https://catalog.ufl.edu/UGRD/academic-regulations/attendance-policies/)

Dropping and Withdrawals

For late course drops and course withdrawals please visit
https://catalog.ufl.edu/UGRD/academic-regulations/dropping-courses-withdrawals/

Incomplete

An incomplete grade may be assigned at the discretion of the instructor as an interim grade for a course in which the student has completed a major portion of the course with a passing grade, been unable to complete course requirements before the end of the term because of extenuating circumstances, and obtained agreement from the instructor and arranged for resolution of the incomplete grade in the next term. Instructors are not required to assign incomplete grades. For complete details please visit CLAS incomplete grade policies and forms.
(https://www.advising.ufl.edu/academicinfo/clas-policiesprocedures/incomplete-grades/)

Accommodating Students with Disabilities

Students requesting accommodation for disabilities must first register with the Dean of Students Office. The Dean of Students will provide documentation to the students who must then provide this documentation to the instructor when requesting information. You must submit this documentation prior to submitting any assignments for which you are requesting accommodation.

Academic Misconduct: Students are held accountable to the UF Honor Code.
(https://sccr.dso.ufl.edu/process/student-conduct-code/)

Evaluations: Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at https://evaluations.ufl.edu. Evaluations are typically open during the last two or three weeks of the semester, but students will be given specific times when they are open. Summary results of these assessments are available to students at
https://evaluations.ufl.edu/results/.

Additional resources: Any additional resources including academic support or information technology can be found at https://www.ufl.edu/about/offices-services/

Approximate weekly breakdown of the material:

Week 1
- Introduction to Statistical Learning

Week 2
- Review of linear regression

Week 3
- Logistic regression
- Linear discriminant analysis

Week 4
- Quadratic discriminant analysis
- Maximal margin classifier

Week 5
- Support vector classifier
- Support vector machines

Week 6
- Cross validation
- The bootstrap

Week 7
- Linear model selection and regularization
- Subset selection
- Ridge regression

Week 8:
- Shrinkage methods
- The lasso
- Dimension reduction methods
- Principle components analysis, Principal components regression
- Partial least squares

Week 9:
- Take home exam

Week 10
- Nonlinear models
- Polynomial regression - regression splines

Week 11
- Smoothing splines
- Local regression
- Generalized additive models

Week 12
- Tree-based methods
- Decision trees

Week 13
- Bagging
- Random forests
- Boosting

Week 14
- Ensemble approaches
- Neural networks

Week 15
- Project presentations