CS 6823: Fundamentals of Machine Learning

Spring 2021

Course Information

Number: 64934
Days: MW
Time: 9:30 - 10:45 am
Credits: 3
Instructor: Emily Bellis, Ph.D.
Room: LSW 236

Office: ABI 202
Office hours: W 10:45-11:45 am or by appointment

email: ebellis@astate.edu

Textbooks and Materials

Required texts: Introduction to Statistical Learning with Applications in R, ISBN 978-1071614174, 2nd edition (2021), by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *A free pdf is available for download at www.StatLearning.com*

Recommended texts:
Elements of Statistical Learning, 12th printing 2017 edition, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *A pdf is available for free download at https://web.stanford.edu/~hastie/ElemStatLearn/*

Additional: Access to a computer with a recent version of R (available for free download at https://cran.r-project.org) and the R Studio integrated development environment (IDE) (https://www.rstudio.com/products/rstudio/). Install R first!

Course Prerequisites

1. One introductory statistics course covering linear regression
2. CS 3113 or "B" or better in CS 5032.

Purpose and Goals

Course Description: Current topics of interest to graduate computer science students. (May be repeated for credit with different subtitle. ONLY six hours with the same course

number will count toward the degree.)

This course is intended to provide an overview to supervised and unsupervised statistical learning techniques commonly used for regression and classification problems in data science. The course focuses on implementing these techniques using the R programming language.

Program Outcomes:
CS 6823 Statistical Learning in Data Science is linked to the following degree-level student learning outcomes for the M.S. in Computer Science: 1) M.S. Computer Science graduate students should have a deeper understanding of the theory and application of algorithms and programming languages. 2) M.S. Computer Science graduate students should have the ability to apply advanced analysis techniques to problem identification and solution in computing applications.

Course Level Learning Outcomes:

Students successfully completing this course will:

- Choose the most suitable approach for analysis of a given dataset among those discussed in the course.
- Defend the choice of approach for a data analysis problem to a broad audience based on its statistical and mathematical underpinnings.
- Gain mastery of the R programming language and familiarity with R Markdown for generation of automated, reproducible, and professional data analysis reports.

Organization

Class sessions are planned to occur for the most synchronously, in person (but see Emergency Instruction Statement). Lectures will include a mixture of powerpoint presentation of the material, active learning activities including peer code reviews, and live coding. Course materials such as rubrics and slides are made available through Blackboard. Students may also track their grades on assignments throughout the course in Blackboard.

Grading

To earn an 'A' for this course, you must:

- Demonstrate mastery on at least 5 'mini-exams' offered during the course AND
- Contribute to at least three code reviews AND
- Earn an A on the Final Project. Final Projects must demonstrate **deep understanding and mastery** of the material covered throughout the course.

To earn a 'B' for this course, you must:

- Demonstrate mastery on at least 5 'mini-exams' offered during the course AND
- Contribute to at least three code reviews AND

- Earn a B or above on the Final Project. Final Projects must demonstrate **good understanding** of the material covered throughout the course.

Students who demonstrate only limited understanding of the course material by the end of the semester (on the mini-exams or on the project) or who do not contribute to peer code reviews will earn a 'C'. Students who engage in academic misconduct (plagiarism or cheating) will earn an 'F'.

The grading policy for this course has been updated as of Fall 2022 after reading Susan Blum's excellent book 'Ungrading: Why Rating Students Undermines Learning'. I appreciate that the lack of a numerical score could be somewhat stressful and will provide a tentative letter grade on your midterm report, so you know where you are headed. All other project reports will be individual feedback (comments) only from myself and from the peer code review. I am keen to know what you think about this new element of the course and if it has helped you focus more on your learning compared to a more 'traditional' policy.

Course Requirements

Mini-exams: In-class mini-exams are quizzes intended to encourage everyone to keep up with the lecture material and the reading, and help both you and me monitor how you are doing. Make-ups will not be given. I will offer at least 6 mini-exams throughout the course, and you will have the opportunity to resubmit the mini-exam, with a written reflection describing why you may have answered a question incorrectly, for full credit.

Reports (5): You will complete five short project reports based on a dataset(s) of your choosing. The goal is to implement and interpret the results of statistical learning techniques covered during the term on a dataset 'from the wild'. *Reports are expected to be the independent work of an individual student and each student must work with a different dataset.* However, you may use the same dataset(s) throughout the semester. You should submit a hard copy of the rendered report that also includes a print out of the .Rmd code (with any mark-up from the peer code review process, and the peer code review itself) attached to the final report as an appendix.

Code Review: Peer code reviews take place in class on days that Reports (see above) are due. Students are expected to bring hard-copies of reports including code to class to facilitate peer reviews; full credit will be given for code reviews that are helpful and constructive (see also Inclusive Environment Policy).

Mid-term project: The mid-term project represents an integration of statistical learning techniques implemented in the first half of the course. You are encouraged to focus on the same dataset and reuse code and results from the Reports (above) and to incorporate improvements from the peer code review and instructor comments. In addition, the mid-term project report will compare the multiple modeling techniques used, defend the most suitable technique for analysis of the chosen dataset, and interpret the results of the best model in the context of a compelling research question.

Final project: The final project represents an integration of statistical learning techniques implemented in the full the course. You are encouraged to focus on the same dataset and

reuse code and results from the Reports (above) and to incorporate improvements from the peer code review and instructor comments. In addition, the final project report will compare the multiple modeling techniques used, defend the most suitable technique for analysis of the chosen dataset, and interpret the results of the best model in the context of a compelling research question. It is not necessary to include all techniques discussed in class; instead, present a comparison of three techniques that are most appropriate to your research question and dataset.

Policies

Late Work: For the most part I will not be accepting late work; if you do not complete a project report on time, you may need to wait until the next due date to receive comments and feedback from me on your project.

Inclusive Environment: I strive for our classroom to be a safe and inclusive learning environment that respects a diversity of thoughts, perspectives and experiences and honors your identities. Please contact me with any concerns that arise and I will do everything in my power to address the situation.

Students with Disabilities: If you have a documented disability, and wish to discuss academic accommodations, please contact me as soon as possible. Students who require academic adjustments in the classroom due to a disability must register with A-State A&AS.

Attendance: Students are generally expected to attend and participate fully in all classes. However, accommodations will be made for students who are not able to attend class, for example if you are quarantined due to COVID-19. **Please do not come to class if you are sick.** Contact the instructor ASAP if you must miss class and would like to know how to make up the missed work.

Academic Misconduct Policy: All project reports and code must be written **in your own words** and be the work of the person seeking academic credit for the course. Any students engaged in plagiarism or cheating will receive a failing grade in the course.

Please see A-State's Academic Integrity Policy in the Student Handbook at http://www.astate.edu/a/student-conduct/student-standards/handbook-home.dot and the CS Department document on Plagiarism in a Programming Context https://wiki.cs.astate.edu/index.php/Plagiarism_in_a_Programming_Context.

Sexual Misconduct Policy: Arkansas State works to provide a safe, productive learning environment. Title IX and university policy prohibit sexual discrimination, which regards sexual misconduct — including harassment, domestic and dating violence, sexual assault, and stalking. Sexual violence can undermine students' academic success.

If you have been sexually assaulted, you can immediately speak with someone at NEARK's Family Crisis Center's 24-hour Sexual Assault Line: (870) 933-9449. Reports to law enforcement can be made to the University Police Department: (870) 972-2093

If you or someone you know has been harassed or assaulted, support can be provided from the Counseling Center and Pack Support. The Student Health Center provides Sexual

Assault resources. Alleged violations can be reported non-confidentially to the Title IX and Institutional Equity office. It provides local, state and national resources for counseling, law enforcement, medical treatment, financial assistance, and legal services.

Title IX Coordinator: title9@astate.edu, Phone: 870-972-2015, Administration Bldg, Room 218A

Notice Concerning the Possibility of Interruption of Instruction Due to an Emergency: While it is the goal of Arkansas State University to offer face-to-face classes for its on-campus programs, the university recognizes that in the event of emergency it may become necessary to shift courses into hybrid or online delivery modes. The recent experience of the COVID-19 pandemic made this necessary; however, the same need to shift could be the product of other natural or civil disasters, and could be for short or extended periods of time. To prepare, this means nearly every course offered will have a component where high-speed, reliable internet access is essential to course success. Other technology such as web cameras or specific software may be required by instructors to facilitate remote instruction (please consult the A-State Internet and Technical Services website for more details). Students are strongly encouraged to secure broadband access they can use for the semester either on or off campus. In the event of the need to change the mode of instruction, A-State will endeavor to keep as many on-campus facilities and support areas open as possible dependent on the circumstances of the emergency.

Please remember, all official notifications are made through your official A-State email account, the university website, and Blackboard Learn. You are responsible for checking your university email to ensure you receive the latest updates regarding this course.

SYLLABUS SUBJECT TO CHANGE AT THE DISCRETION OF THE INSTRUCTOR

Table 1: CS 6823 Schedule

Week	Date	Topic	Due	Reading
1	8/24/22	Intro to Statistical Learning		Chp. 1-2
2	8/29/22	R & R Markdown		
	8/31/22			
3	**9/5/22**	**No Class; Labor Day**		
	9/7/22	Regression		Chp. 3
4	9/12/22			
	9/14/22	Classification		Chp. 4
5	**9/19/22**		**Code Review (Report 1)**	
	9/21/22			
6	9/26/22	Resampling		Chp. 5
	9/28/22			
7	10/3/22	Alternatives to Least Squares		Chp. 6
	10/5/22			
8	**10/10/22**		**Code Review (Report 2)**	
	10/12/22			
9	**10/17/22**	**Beyond Linearity**	**MIDTERM PROJECT (Report 3)**	**Chp.7**
	10/19/22			
10	10/24/22			
	10/26/22	Decision Trees		Chp. 8
11	10/31/22			
	11/2/22			
12	11/7/22	Unsupervised Learning		Chp. 12
	11/9/22		**Code Review (Report 4)**	
13	11/14/22			
	11/16/22	SVMs		Chp. 9
14	**11/21/22**	**No Class; Fall Break**		
	11/23/22	**No Class; Fall Break**		
15	11/28/22	Deep Learning		Chp. 10
	11/30/22			
16	**12/5/22**	**Last day of class**	**Code Review (Report 5)**	
17			**FINAL PROJECT**	