Statistics 6289 - Spring 2022
Statistical Machine Learning
6:10 - 8:40pm Thursday – Funger Hall, Room 222

Instructor:	Dr. Darcy Steeg Morris
Office Hours:	after class or by appointment
E-mail:	dsteeg@gwu.edu
TA:	Zhou Yang
Office Hours:	Monday 2-3pm and Wednesday 3-4pm
E-mail:	zhou_yang@email.gwu.edu

Course Description: A study of statistical and theoretical foundations of modern statistical machine learning. Topics covered include: resampling methods (cross-validation and bootstrap), model evaluation and variable selection, maximal margin classifiers and support vector machines, tree-based methods for regression and classification, ensemble methods, mixture models, and unsupervised learning,

Prerequisite(s): Advanced knowledge of R programming. Graduate level mathematical statistics courses at the level of 6201 and 6202. Graduate level linear model course at the level of 6214.

Required Text:

The Elements of Statistical Learning, 2nd Edition by Hastie, Tibshirani & Friedman.
    Available online: www.web.stanford.edu/~hastie/ElemStatLearn.

Recommended Texts:

Pattern Recognition and Machine Learning (2006) by Bishop.*

An Introduction to Statistical Learning, 2nd Edition by James, Witten, Hastie & Tibshirani.
    Available online: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf.
    Videos and notes: https://tdg5.github.io/stats-learning-notes/.

Learning Outcomes: At the completion of this course students will be able to: (1) formulate the statistical/probabilistic structure of common machine learning techniques, (2) understand theoretical aspects of the design and evaluation of statistical learning methods/algorithms, (3) identify appropriate statistical learning methods for different types of data, and (4) program, evaluate, and interpret statistical learning methods and algorithms in R.

Grading:

Homework Assignments	65%
Homework Presentation	5%
Special Topic Lecture	30%

- *Homework Assignments*: Homework will be assigned every other week. Late assignments will not be accepted. Students are expected to write up their own solutions to all assigned problems. All assignments are equally weighted.

- *Homework Presentation*: Each week 4-5 students will be assigned to present some portion of their homework solutions to the class. The part of the homework to be presented will not be known in advance. Students will be graded on their communication and clarity in addition to the validity of their solution. The assigned order will be alphabetical – the first 4-5 students by alphabetical order present on Homework #1, the next 4-5 present on Homework #2, etc.

- *Special Topic Lecture:* Students will present a special topic from one of the textbooks. Each group of students are responsible for teaching classmates their chosen material to include a data analysis example, and submitting a write-up of their special topic.

- Students are expected to attend all lectures during class time (2.5 hours a week) either in-person or virtually.

- Students are expected to spend a minimum of 10 hours a week on out-of-class/independent learning.

Mode of Instruction:

Lectures will begin as virtual class meetings on Blackboard Collaborate Ultra (BBCU). Lectures will be in-person starting the second week of class, but will be synchronously streamed online through BBCU for students who have university approval to attend virtually. You are permitted to attend a given lecture virtually on a case-by-case basis with approval from me in advance. You are expected to attend all lectures in-person or virtually (if approved), but lectures will be recorded and posted in case virtual attendees experience technical difficulties and/or internet connectivity issues.

Blackboard:

All students are required to register for the course in Blackboard, the GWU web-based instructional resource. Course information and materials, including homework assignments and grades will be posted there. It is the students responsibility to check the Blackboard website frequently for up-to-date information.

Use of Electronic Course Materials and Class Recordings:

Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Please contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Please contact Disability Support Services at www.disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Computing:

We will use R for statistical computing. Please download the R software online at: www.r-project.org and www.rstudio.com.

Academic Integrity Code:

Academic Integrity is an integral part of the educational process, and GW takes these matters very seriously. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and in other ways outlined in the Code of Academic Integrity. Students accused of academic integrity violations should contact the Office of Academic Integrity to learn more about their rights and options in the process. Outcomes can range from failure of assignment to expulsion from the University, including a transcript notation. The Office of Academic Integrity maintains a permanent record of the violation.

More information is available from the Office of Academic Integrity at www.studentconduct.gwu.edu/code-academic-integrity. The Universitys "Guide of Academic Integrity in Online Learning Environments" is available at www.studentconduct.gwu.edu/guide-academic-integrity-online-learning-environments. Contact information: rights@gwu.edu or 202-994-6757.

Safety and Security:

- In an emergency: call GWPD 202-994-6111 or 911.

- For situation-specific actions: review the Emergency Response Handbook at `www.safety.gwu.edu/emergency-response-handbook`.

- In an active violence situation: Get Out, Hide Out, or Take Out. See `www.go.gwu.edu/shooterprep`.

- Stay informed: `www.safety.gwu.edu/stay-informed`.

University Policy on Observance of Religious Holidays:

In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance. For details and policy, see "Religious Holidays" at `www.provost.gwu.edu/policies-procedures-and-guidelines`.

Support for Students Outside the Classroom:

Virtual Academic Support

A full range of academic support is offered virtually in Fall 2020. See `https://coronavirus.gwu.edu/top-faqs` for updates. Tutoring and course review sessions are offered through Academic Commons in an online format. See `https://academiccommons.gwu.edu/tutoring`. Writing and research consultations are available online. See `https://academiccommons.gwu.edu/writing-research-help`. Coaching, offered through the Office of Student Success, is available in a virtual format. See `https://studentsuccess.gwu.edu/academic-program-support`. Academic Commons offers several short videos addressing different virtual learning strategies for the unique circumstances of the Fall 2020 semester. See `https://academiccommons.gwu.edu/study-skills`. They also offer a variety of live virtual workshops to equip students with the tools they need to succeed in a virtual environment. See `www.tinyurl.com/gw-virtual-learning`.

Writing Center

GWs Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies to write independently in academic and public settings. Appointments can be booked online. See `https://gwu.mywconline`.

Academic Commons

Academic Commons provides tutoring and other academic support resources to students in many courses. Students can schedule virtual one-on-one appointments or attend virtual drop-in sessions. Students may schedule an appointment, review the tutoring schedule, access other academic support resources, or obtain assistance at `https://academiccommons.gwu.edu`.

Disability Support Services (DSS): 202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services to establish eligibility and to coordinate reasonable accommodations. For additional information see: `www.disabilitysupport.gwu.edu`.

Counseling and Psychological Services : 202-994-5300

GWs Colonial Health Center offers counseling and psychological services, supporting mental health and personal development by collaborating directly with students to overcome challenges and difficulties that may interfere with academic, emotional, and personal success. For additional information see: `www.healthcenter.gwu.edu/counseling-and-psychological-services`.

Course Schedule:

The weekly coverage and number of homework assignments (and due dates) may change as they depend on the progress of the class. The instructor reserves the right to make changes to this course schedule as necessary.

Week	Content
Week 1 (1/13)	• Introduction and Overview (Virtual on BBCU) • *Reading assignment*: Chapter 1, 2.1-2.3, 2.9
Week 2 (1/20)	• Statistical Decision Theory and Function Approximation (Begin In-Person) • *Reading assignment*: Chapter 2.4, 2.6-2.7
Week 3 (1/27)	• Model Evaluation and Resampling Methods • *Reading assignment*: Chapter 7.1-7.7, 7.10-7.11 • **Homework #1 Due**
Week 4 (2/3)	• Likelihood and Bayesian Estimation/Inference • *Reading assignment*: Chapter 8.1-8.4, 8.6
Week 5 (2/10)	• Linear Regression and Regularization/Shrinkage • *Reading assignment*: Chapter 3.1-3.5, 3.6-3.9 • **Homework #2 Due**
Week 6 (2/17)	• Splines and Basis Expansion • *Reading assignment*: Chapter 5.1-5.2, 5.4-5.5, 9.1
Week 7 (2/24)	• Logistic Regression and Discriminant Analysis • *Reading assignment*: Chapter 4.1-4.4 • **Homework #3 Due**
Week 8 (3/3)	• Maximal Margin Classifier, Soft Margin Classifier, and Support Vector Machines • *Reading assignment*: Chapter 4.5, 12.1-12.3
Week 9 (3/10)	• Decision Trees, Bagging/Boosting/Model Averaging, and Random Forests • *Reading assignment*: Chapter 8.7-8.8, 9.2, 15.1-15.4 • **Homework #4 Due**
Week 10 (3/17)	• *Spring Break*
Week 11 (3/24)	• Mixture Models and EM Algorithm • *Reading assignment*: Bishop Chapter 9.2-9.4 • **Homework #5 Due**
Week 12 (3/31)	• Undirected Graphical Models and Markov Models • *Reading assignment*: Chapter 17.1-17.4
Week 13 (4/7)	• High-Dimensional Data, Dimension Reduction, and Cluster Analysis • *Reading assignment*: Chapter 14.1, 14.3, 18.1-18.4 • **Homework #6 Due**
Week 14 (4/14)	• **Student Lectures**
Week 15 (4/21)	• **Student Lectures**