DATA 5322 Statistical Machine Learning II
Spring Quarter 2024

Professor's Information:
Dr. Ariana Mendible
Pronouns: she/her/hers
Email: mendible@seattleu.edu
Phone: (206) 296-5864
Office: BANN 428

Course Information:
Credits: 3
Location: LEML 122 (Boeing Room)
Lecture: Monday 6:00-8:40 pm
Office Hours: TBD by survey
and by appointment

Course Description

This course expands upon Statistical Learning I to introduce a number of additional common tools in statistical and machine learning beyond the linear modeling framework. Students will learn classification methods. Standard approaches to model regularization and dimension reduction will be covered. Students will also be introduced to unsupervised learning/clustering methods.

Prerequisites DATA 5321 - Statistical Machine Learning I

Textbook An Introduction to Statistical Learning with Applications in Python/R, by James, Witten, Hastie, & Tibshirani. Available for free online.

Course Topics Tree-based methods, support vector machines, neural networks, regularization and dimension reduction, unsupervised learning

Learning Outcomes

On successful completion of this course (i.e. by passing this course), you will

1. Be able to use machine learning techniques in the context of data science methodology;
2. Apply statistical models, tree-based models, support vector machines, and unsupervised learning models;
3. Develop models with parameter regularization and dimension reduction;
4. Select an appropriate model for a problem, including understanding the differences between models based on categorical versus numerical responses;
5. Assess the performance of a statistical model;
6. Recognize when unsupervised methods are an appropriate modeling tool.

Program Learning Outcomes Addressed

(1) Apply appropriate analytical and computational methods to solve real-world problems effectively. (2) Communicate technical information effectively to a specific audience via speech, writing, and data visualization. (3) Exhibit constructive and inclusive collaboration and teamwork skills.

Graduate Learning Outcomes Addressed

(1) Demonstrate mastery of competencies required in their profession or field. (2) Demonstrate effective communication in speech and in writing. (3) Exhibit professional integrity, ethical leadership, and effective collaboration skills. (4) Recognize and address moral and ethical challenges within their professions as informed by the Jesuit Catholic tradition. (5) Develop a professional perspective focused on life-long learning that is informed by the knowledge and skills of their graduate education.

Evaluation and Assignments

This course is designed to train students in practical applications for careers in data science. To that end, students will be evaluated on a mix of computational and communication skills. There will be no timed exams. Active participation in class is expected and will be evaluated regularly.

Weekly Worksheets (20%)

Each class day, a exploratory worksheet will be posted on Canvas. Students will work together in class to explore the topic at hand. After class, students will complete outstanding work, polish the documents with clean code, verbal answers to response questions, plots/outputs, and comments, and will submit the work on Canvas. The goal of these assignments is to provide low-stakes feedback on learning and ensure timely course participation. They are due by the start of the next class period, typically one week out. These will be graded for completion and effort. No late submisisons will be accepted. At the end of the quarter, the lowest worksheet grade will be dropped.

Written Homework (50%)

Written homework will be assigned with each textbook chapter, approximately every 2 weeks, and will explore the full data science methodology of collecting, processing, and analyzing data as well as communicating results. Typically, these assignments will be exploratory and are meant to reflect real-world problems. They are built for students to showcase mastery of course content, but will be based on real data and may require additional resources, data cleaning, or other tools that are not directly presented in class. **As such, it is expected that these assignments will take longer to complete.** Each written homework will have two deliverables: (1) your code and (2) a written report of your work. The reports should be written to explain the concept/theory, your methodology, and your findings. They should be professional and polished and will be graded on the presentation and communication (neat, complete, clear language, sufficient plots that are correctly labeled and discussed, etc). Expectations for written communication are high– students can seek extra support at the Writing Center and the English-Language Learning Center. A report format template, example report, and grading rubric will be posted on Canvas to give students a clear idea of what this should look like. Written homework will be due one week after it is assigned. Revisions are encouraged and will be due one week after feedback is recieved.

Summary Map (5%)

Upon completing this course, students will have learned a variety of methods and approaches to machine learning, each with accompanying advantages and disadvantages, appropriate uses, and parameter choices. In order to synthesize the information from this course (and Statistical Machine Learning I), students will develop a summary document of this information. This should be worked on throughout the quarter to synthesize knowledge at the end of a module. It will also serve as a quarter-end study guide and as a reference for the future (preparing for job interviews, for example).

Final Project (25%)

Students will work in small groups (3-4) to create and solve a data science problem of their choosing. It should showcase their knowledge of machine learning concepts and solve a problem that is compelling to the group. The final project will be presented in advance of course completion so that students have ample time to find a topic, gather data, complete the project, and prepare their deliverables. Deliverables for this project are (1) a written report, more in-depth but in a similar format to homeworks, and (2) an oral presentation. During finals week, groups will give a 10-15

minute presentation to the class and community about their topic of study. All group members are expected to participate equally in all aspects of the project. Exceptions for illness or other extenuating circumstances should be communicated and will be accommodated as best as possible.

Classroom Norms

It is expected that all students participate fully in class activities. This includes arriving in class on time, actively engaging with others during group work, and sharing ideas and questions openly. My foremost goal is student learning. If there is anything I can do to better facilitate your learning (e.g. assignment extensions, extra help, changing some of the course structure), please feel free to let me know.

Academic Integrity

While working together outside of class and using a variety of resources is **strongly encouraged**, it is considered cheating to simply rephrase work generated by another student or outside resource. Students are encouraged to learn from one another and all of the other resources you have at your disposal, but work that is handed in must ultimately reflect a student's own understanding of the course material. Cheating and plagiarism are defined in the Academic Integrity Policy. If the professor suspects that a student has cheated or plagiarized, she will follow the procedures outlined in the University's policy. In cases of cheating or plagiarism, the assignment or exam will receive a zero and this score will be included in the student's final grade for the course. Seattle University's Academic Integrity Tutorial can be found here.

General Course and University Policies

Support for Students with Disabilities: If a student has, or think they may have, a disability (including an "invisible disability" such as a learning disability, a chronic health problem, or a mental health condition) that interferes with their performance as a student in this class, they are encouraged to arrange support services and/or accommodations through Disability Services staff. They can be reached at (206) 296-5740. Disability-based adjustments to course expectations can be arranged only through this process.

Religious Accommodations: It is the policy of Seattle University to reasonably accommodate students who, due to the observance of religious holidays, expect to be absent or endure a significant hardship during certain days of their academic course or program. Please see, the Policy on Religious Accommodations for Students.

Community and Inclusivity Seattle University and the Department of Mathematics are committed to creating and sustaining an inclusive culture that values diversity and works for equity in opportunity and outcomes. Diversity is a core value we espouse as part of our mission. We respect our students' identities and we strive to create a learning environment where every student feels welcomed and valued. We ask for your help in fostering a welcoming and open environment, treating others with respect, and collaborating toward equity. Please refer to the Student Code of Conduct and to the Office for Diversity and Inclusion for more information and guidance.

Campus Climate Incident Reporting & Response Protocol Seattle U has adopted a Campus Climate Incident Reporting & Response Protocol to support our commitment to providing an inclusive and nondiscriminatory campus community. If you have seen, heard, or experienced a harmful incident on the basis of one or more of your or another individual's actual or perceived identities, you may report that incident here.