Syllabus

Math 457 Introduction to Statistical Learning. Fall 2024.

Binghamton University

- Instructor: Vladislav Kargin
- Office: WH-136
- Meeting time and location: MWF – 11:20-12:50 – CW 213
- Office hours: MWF – 2 - 3pm

This course is a 4-credit course, which means that in addition to the scheduled lectures/discussions, students are expected to do at least 9.5 hours of course-related work each week during the semester. This includes things like: completing assigned readings, participating in lab sessions, studying for tests and examinations, preparing written assignments, completing internship or clinical placement requirements, and other tasks that must be completed to earn credit in the course.

Prerequisite

- Scientific programming in a language such as R, Matlab, or Python.
- Linear regression and statistical inference
- Matrix algebra, preferably including orthogonality, eigenvalues and eigenvectors, and singular value decomposition.

Description

This course is a survey of statistical learning methods. It will cover major statistical learning methods and concepts for both supervised and unsupervised learning. Topics covered include regression methods with sparsity or other regularizations, model selection, introduction to classification, including discriminant analysis, logistic regression, support vector machines, and kernel methods, nonlinear methods, clustering, decision trees, random forest, boosting and ensemble learning, neural networks, survival analysis.

Learning Outcomes

Students will learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms. Students completing this course should be able to

- process and visualize different data types,
- apply basic statistical learning methods to build predictive models or perform exploratory analysis
- have basic understanding of the underlying mechanism of predictive models and evaluate and interpret such models,
- properly tune, select and validate statistical learning models,
- use analytical tools and software widely used in practice,
- work both independently and in a team to solve problems, and
- learn to present and communicate the findings effectively.

Text

- James, Witten, Hastie and Tibshirani, 2021. “An Introduction to Statistical Learning with Applications in Python.” The Book Home Page is at “https://www.statlearning.com”. A pdf file can be downloaded from this page.

Online resources

There is an online course taught by the book's author available at YouTube course.

Software

We will use Python for this class. We will use Google Colab. Optionally, you can install Anaconda and Jupiter on your computer to run Jupiter notebooks with Python code locally. See some instructions here: Installation instructions.

Piazza

We will use Piazza (“http://piazza.com/”) for communication. All announcements will be sent to the class using Piazza. The signup is at this link: https://piazza.com/binghamton/fall2024/math457.

Mycourses

Mycourses (“http://mycourses.binghamton.edu”) will only be used occasionally.

Homework / Quizzes Policies

The Homework will have several components:

1. Several Assignments at Datacamp for learning python (worth 12.5% of HW)
2. Computational Assignments as Google Colab Notebooks (worth 37.5% of HW)
3. Assignments at Gradescope /(or possibly in-class Quizzes) (worth 50% of HW)

We will use Datacamp, Google Classrooms (Class code: rm3nfmh) and Gradescope (“https://www.gradescope.com/”) to submit and grade homework. Homework may be discussed with classmates but must be written and submitted individually. ChatGPT and similar AI tools can be used for homework. They are not allowed during exams.

If in-class quizzes are used, the policy is that two lowest or missing grades are dropped from total score calculation.

Exam

There will be a midterm and a final exam focusing on the theoretical part of the course. Final is cumulative.

Project

A group project will be assigned to each student (2 - 3 students in a group, 4 students are not allowed without a strong justification). Successful completion of the project includes an preliminary report, a presentation and a final report.

Grading

- Homework (30%)
- Midterm exam (20%)
- Project (30%)
- Final exam (20%)

Tentative schedule

Midterm	Sep 27
Project Proposal	due Nov 1
Preliminary report	due Nov 13
Project presentations	Dec 2 - Dec 4
Final Report	due Dec 6
Final Exam	As scheduled by the University, —