Syllabus

Math 457 Introduction to Statistical Learning. Fall 2021.

Binghamton University

- Instructor: Vladislav Kargin
- Office: WH-136
- Meeting time and location: MWF 8:00 - 9:30 am at OH-G102.
- Office hours: MWF 9:45 - 10:30 (in person, office WH136), Tue 4:00PM - 5:00PM (via Zoom, ID 949 5616 9870), or by appointment

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 its 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, deep learning

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.

Recommended Texts

- James, Witten, Hastie and Tibshirani, 2021. “An Introduction to Statistical Learning with Applications in R.2nd edition” The Book Home Page is at “http://www-bcf.usc.edu/~gareth/ISL/index.html”. A pdf file can be downloaded from this page.

Online resources

There is a course taught by Hastie and Tibshirani using the first edition of their book. This Course is available at edx. The course is not free, however the videos and some other resources are available to auditors. The videos can also be obtained at this website through playlist links.

Software

We will use R and R Markdown for this class. The IDE for R, RStudio can be downloaded from here.

Piazza

We will use Piazza (“http://piazza.com/”) for communication. All announcements will be sent to the class using Piazza.

Gradescope

We will use Gradescope (“https://www.gradescope.com/”) to submit and grade homework. This will allow the instructor to efficient grade all the work and give feedback in a timely manner.

Mycourses

Mycourses (“http://mycourses.binghamton.edu”) will only be used occasionally for recording grades on assignments and exams and for distributing solutions.

Homework Policy

Homework will be assigned approximately bi-weekly. It is expected that homework is prepared using R Markdown or LaTeX. Handwritten homework is not accepted. There will be a deduction of 15% of the grade for each day homework assignment is late (the final grade for a late homework that is N days late will be 0.85^N times the real grade). Homeworks may be discussed with classmates but must be written and submitted individually.

Midterm Exam

A midterm exam focusing on the theoretical part of the course will be administered in November.

Project

A group project will be assigned to each student (2 - 4 students in a group). Successful completion of the project includes an initial report, a presentation and a final report.

Grading

- Homework (40%): homework is assigned biweekly.
- Midterm exam (30%):
- Project (30%)
- Lecture attendance and participation – possible bonus of up to 3% according to the instructor judgement.

Tentative schedule

Midterm	Nov 22
Project Proposal	due Nov 24
Preliminary report	due Dec 3
Project presentations	December 6, 8, 10
Final Report	due December 13