Syllabus

DATA SCI 415

Winter 2025

1 Teaching Team

Instructor: Snigdha Panigrahi

- Lecture: Mon, Wed 2:30-4pm, UMMA Aud
- Office hour: Mon 9:30-10:30am, 451 West Hall

GSI: Alexander Kagan, Samuel Rosenberg

- Labs: Tues 10:00-11:30, 11:30-1.00
- Office hour: TBA

2 Material

- Textbook: James, Witten, Hastie and Tibshirani (2015) An Introduction to Statistical Learning. Springer.
- Lecture notes; Use the textbook as a supplement to the Lectures
- Lab Assignments, Homework Problem Sets
- Software: Python

3 Course outline

This course will cover data mining and various tools for conducting exploratory data analysis, and supervised as well as unsupervised machine learning. It will include both linear and non linear methods for regression and classification, and methods for dimension reduction. The course will cover statistical concepts and computational issues. Students can hope to gain knowledge about learning algorithms, statistical computing, and core statistical concepts driving these algorithms.

Here is a brief list of topics that we will cover in this class.

- Nature of data and exploratory data analysis
- Linear regression: Least squares and shrinkage
- Classification: Logistic regression, Bayes rule, LDA, QDA
- Non-linear methods for regression and classification: Polynomial, Splines, GAMs, Trees, Ensemble methods, Neural Networks
- Unsupervised learning: Clustering, PCA

4 Exams and Assessment

- 5 (long) homework problem sets, 2 exams, and regular lab assignments
- First exam: Oct 22, 2:30-4pm, UMMA Aud
- Second exam: Dec 10, 4-6pm, UMMA Aud
- Homework 10%, labs 20%, first exam 30%, second exam 40%

More on the different components of this class.

- A random subset of the 5 homework problems will be graded. Similar questions might also appear on your graded lab assignments!!

  Think of the homework as extra practice for your labs—completing it on your own will help you perform better on the graded lab assignments.
- Lab assignments will be done in pairs! There will be 12 Lab assignments, out of which the best 10 will count towards your final grade.
- Late homework or lab submissions will not be accepted. Email requests for exceptions will not be considered.

Homework.

- Homework will be submitted electronically through Canvas as a pdf
- Posted and due on Fridays

Exams.

- Exams are closed book and do not involve a computer
- You are allowed to bring one standard size sheet of paper, writing whatever you want on both sides, and a calculator
- Exams will not involve any coding, though they may require understanding code snippets
- 1 Practice Exam will be provided before each Exam

Use of Gen-AI.

In principle you may submit AI-generated code, or code that is based on or derived from AI-generated code, as long as this use is properly documented in the comments: you need to include the prompt and the significant parts of the response. AI tools may help you avoid syntax errors, but there is no guarantee that the generated code is correct. It is your responsibility to identify errors in program logic through comprehensive, documented testing. Please keep in mind: generated code, even if syntactically correct, may have significant scope for improvement, in particular regarding separation of concerns and avoiding repetitions.

5 Other components of this class

Labs

- Labs will be instructed by your GSIs

- First Lab starts next week (Week starting Sept 1) : No lab tomorrow!
- More on labs and assignments: next week (you'll have a warm-up)

Discussion beyond lectures and OHs.

- Piazza policy: GSIs will rotate and will state this policy in the first Lab.