BE 559: Foundations of Biomedical Data Science and Machine Learning (Spring 2025)

Instructors: Michael Economo (mne@bu.edu) & Brian DePasquale (bddepasq@bu.edu)
TA: Isaac Boyd (ipboyd@bu.edu)

Class: T/Th 1:30-3:15
Location: CDS 262

Recitation: F 2:30-3:35
Programming help: F 3:35-4:25
Location: ERB 203

Office Hours: By appointment. 24 Cummington Mall, Room 201 (Economo) & 24 Cummington Mall, Room 203 (DePasquale)
Course documents: On Blackboard (https://learn.bu.edu)

Course description: This course will cover conceptual and practical aspects of data science and introductory machine learning for biomedical engineers. This course serves as a foundational course in data analytics for BME Ph.D. students. It is designed to follow a graduate-level introductory programming course and will prepare students for graduate-level courses and research focused on more advanced applications of machine learning and data science. This course will cover the theory and practical applications of hypothesis testing, model fitting and parameter estimation, classification, clustering, dimensionality reduction, and machine learning.

Course Goals: This course takes a practical approach to the analysis of biomedical data. By doing so, three goals are strived for. First, students will become familiar with the necessary theoretical background of different analysis methods, empowering them to understand why certain methods are appropriate in certain contexts and why others are not. Second, students will acquire practical, hands-on skills necessary for analyzing biomedical data including data management, algorithm development, and proper codebase development. These skills will prepare students for independent research projects both within academic research and within industry. Third, students will learn how to interpret, visualize, and summarize the results of analysis, once completed. Applying analysis methods is only half the challenge of scientific discovery. The third goal of this course is to train students to collect the results of scientific analysis into a format that allows scientific discoveries to be shared with and understood by other researchers.

Prerequisites: BE 601 (Linear Algebra), or equivalent familiarity with linear algebra (e.g., linear systems of equations, eigendecomposition, bases, orthogonality, matrix decomposition, etc.), and BE 604 (Statistics & Numerical Methods) or equivalent familiarity with probability and statistics (e.g., random variables, conditional and marginal probability, Bayes' rule, etc.). ENG BE 500 (Programming Fundamentals for Biomedical Engineering Data Analysis with Python) or equivalent programming experience.

Grading & exercises: Student grades will largely be determined by weekly coding/analysis assignments that will together comprise 50% of the course grade. The remaining 50% will be based on a final exam.

Programming exercises are to be completed in Python using Google Colab notebooks. Course assignments will be distributed by Blackboard and solutions are to be submitted (in .ipynb format) electronically via GradeScope on or before the assigned due date. Collaboration on the homework is allowed (and encouraged!) but your submitted work must be your own (see below). The best way to fill out answers in google colab is to create a copy of the provided homework template (File > Save Copy in Drive), and then use this copy to populate your answers to the questions.

Collaboration/plagiarism: Plagiarized assignments will be assigned a zero grade. Students found to have plagiarized assignments or provided their assignments to others may receive a failing grade in the course and/or a referral to the College of Engineering Academic Conduct Committee. Below is clarification about what IS and what IS NOT acceptable. If you are uncertain if actions are acceptable or unacceptable, please ask us!

DO: Verbally exchange tips and insights about coding, concepts from class, or anything else that is unclear. This could include looking at short segments of code together, discussing why certain solutions work, and debugging.
DON'T: ***Electronically* share code and other work with other students**. Copy and paste work of another student (even with variable names, etc. altered!). Copy and paste code from other sources (e.g StackOverflow) that you do not fully understand.

Policy for late work: For all homework assignments, scores will be reduced by 25% if turned in within one week of the scheduled due date. Scores will be reduced by 50% if turned in more than one week late.

Final Exam: The final exam will be held during the course's scheduled exam slot (May 6th at noon). The final exam will consist of short answer questions aimed at assessing students conceptual comprehension of covered topics.

Recitation and programming help: Recitations will cover questions from lecture, example problems from the week's material, and introduce homework material. Then, there will be an optional programming help session where we will review basic programming concepts and provide support for those learning advanced programming packages (e.g., Tensorflow, PyTorch, SciPy, etc.).

Textbook: Pattern Recognition and Machine Learning (Bishop; PDF) and various other texts (all freely available in PDF form).

Course topics (Spring 2025)

Module & Lectures	Topics
Module 1 - Hypothesis testing	• Intro to Python • Review of probability • Review of statistics • Introduction to Hypothesis testing • Parametric vs. non-parametric statistics • Multiple comparisons • Resampling-based statistical methods • Advanced methods
Module 2 - Regression and model fitting	• Fitting models to data • Likelihood functions/Maximum likelihood solutions • Basics of optimization • Linear regression • Python functions • Logistic regression • Regularization • Bayes' Rule • Bayesian perspective link to regularization • Model validation, AIC, generalizability, cross-validation
Module 3 - Classification and clustering	• Types of errors • Receiver-operating characteristic (ROC) curves • Linear discriminant analysis (LDA) • Bayesian classifiers • Expectation maximization (EM) • K-means • K-nearest neighbors (K-NN) • Hierarchical clustering • Distance measures • Advanced methods
Module 4 - Dimensionality reduction	• Linear algebra review (bases, eigendecomposition, etc.) • Principal components analysis • LDA (higher dimensional) • Revisiting regression in light of PCA (PLS, PCR, Ridge) • Revisiting expectation maximization • Factor analysis • Advanced methods
Module 5 - Neural networks	• Gradient descent/re-introduce optimization • Automatic differentiation • Backpropagation algorithm • Programming neural networks

	• Neural network loss functions and information theory

Tentative course schedule (Spring 2025)

Week	Date	Tues lecture	Thurs lecture	Notes	Lecturer
Week 1	1/21, 1/23	1.1	1.2		Mike
Week 2	1/28, 1/30	1.3	1.4		Mike
Week3	2/4, 2/6	1.5	1.6		Mike
Week 4	2/11, 2/13	2.1	2.2	**2/11:** HW1 due @ noon	Brian
Week 5	2/20	—	2.3		Brian
Week 6	2/25, 2/27	2.4	2.5		Brian
Week 7	3/4, 3/6	3.1	3.2	**3/4**: HW2 due @ noon	Mike
Week 8				**Spring Break**	
Week 9	3/18, 3/20	3.3	3.4		Mike
Week 10	3/25, 3/27	3.5	4.1		Mike/Brian
Week 11	4/1, 4/3	4.2	4.3	**4/1**: HW3 due @ noon	Brian
Week 12	4/8, 4/10	4.4	4.5		Brian
Week 13	4/15, 4/17	4.6	4.7		Brian
Week 14	4/22, 4/24	5.1	5.2	**4/22:** HW4 due @ noon	Brian
Week 15	4/29, 5/1	5.3	review		Brian
Week 16	5/6	Final Exam			