BIOST 2155 – Introductory Statistical Learning for Health Sciences – Fall 2025
Credit Hours: 2 credits

Last updated: 08/27/2025

1. Instructor:

Soumik Purkayastha, PhD, Assistant Professor of Biostatistics and Health Data Science
E-mail: soumik@pitt.edu
Phone: (412) 624-0336

2. Times and Location:

Lecture:
    Time: Friday 9:00-10:50 AM (8/29/2025–12/05/2025)
    Location: A622 Public Health (in-person)

Office hours:
    Time: Friday 11:00–11:40 AM (8/29/2025–12/05/2025) or by appointment (virtual or in-person).
    Location: A740 Public Health (in-person).

Teaching Assistants:

Jessica Shao (yus221@pitt.edu)	Hao Wang (haw291@pitt.edu)
Office Hour Time: Wednesday 12:00 – 1.00PM	Office Hour Time: Monday 11:30 – 12.30PM
Office Hour Location: A724A	Office Hour Location: A724A

3. Course Description

This 2-credit course is a graduate-level course to introduce basic concepts and methods for statistical learning with an emphasis on modern health science applications. The syllabus includes linear regression with regularization, supervised machine learning, unsupervised clustering, and dimension reduction. The target audience will be second-year Biostatistics master students or early PhD students with interests in statistical learning techniques for health science data. Students will be trained with hands-on materials to understand the methods, implement the algorithms, and interpret results in real applications.

4. Prerequisites

BIOST 2141, 2142, and 3011. Students are expected to have programming experiences in R or in some low-level languages such as C, C++, Java and Fortran. Prerequisites can be met with other relevant courses/experience with consent from the instructor.

5. Learning Objectives

At the conclusion of this course, the students should be able to:
- Explain the motivation and insights behind statistical learning methods covered in class.
- Demonstrate basic theory proof for selected fundamental concepts and methods.
- Implement methods to real datasets in health sciences.

- Produce an analysis pipeline and interpret the results in a real application.

6. Textbook

Relevant lecture notes, slides and reading materials will be made available on Canvas. The following textbook is highly recommended and will be used for most lectures.
- James, Witten, Hastie and Tibshirani. An Introduction of Statistical Learning: with Applications in R, Springer, 2013. [link]

The following two books are optional references. Free copies are available online.
- Hastie, Tibshirani, and Friedman. The Elements of Statistical Learning. Springer, 2009. [link]
- Prince. Understanding Deep Learning. MIT Press, 2024. [link]
- Bishop. Pattern Recognition and Machine Learning. Springer New York, 2006. [link]

7. Canvas Instruction

Canvas (https://canvas.pitt.edu/courses/324229) will serve as the main archive of lecture slides, lecture notes, homework assignments, and other materials. Announcements concerning course logistics will also be sent and posted on the webpage.

8. Computing

We will use R as our main programming language. You can consider using RStudio and its R Markdown feature to generate homework reports. Template will be provided.

9. Grading Scale and Student Performance Evaluation (Assessments and Weights)

Course grades will be based on a weighted average of
- Homework assignments 75% (Three homework assignments; each 25%)
- Final project 25%: Final project presentation (25%) (analysis of a real dataset or review of a methodological paper)

The cut-offs for computing letter grades will be: A, 100%-90%; B, 89%-80%; C, 79%-70%; D, 69%-60%; and F, <60%. Plus-minus grades will be assigned by dividing the respective intervals into thirds. Discussions of homework assignments among students are allowed. But each student must independently write and implement their own solution. Solutions will be graded on both correctness and clarity. Credits will be given for trying to identify the gaps in argument if complete solution cannot be derived. Cheating and plagiarism is strictly not allowed and may be reported to the university. See the University of Pittsburgh's Policy on Academic Integrity at http://www.provost.pitt.edu/info/ai1.html

Late homework will be discounted accordingly: 80% if 0-1 days past due, 50% if 1-2 days, 0% if later than 2 days. Exceptions can be made in case of sickness or emergency, under which the instructor must be informed in advance.

Homework
There will be 3 homework assignments (mixture of theory proof, simulation and real data application using R). Students will turn-in an electronic copy via Canvas. *Use white space and include clear comments to make code readable.*

Final projects
By 10/03, students should determine a selected dataset or paper and submit a one-page proposal detailing planned project work. On the final day of class (12/05), everyone will have 5 minutes to present their individual work and summarize their findings.

10. Course Schedule

Preliminaries		
08/29	**Introduction**: supervised learning (regression and classification); unsupervised learning (clustering and dimensionality reduction).	ISLR 2
09/05	**Training statistical learning models**: under- and over-fitting; training, validation, and testing; resampling techniques.	ISLR 2.2, 5
Supervised learning I: Regression		
09/12	**Regression (I):** regularization using subset selection; shrinkage.	ISLR 3, 6
09/19	**Regression (II):** non-linear regression; splines; kernel trick.	ISLR 7
Supervised learning II: Classification		
09/26	**Classification (I):** basics; generative models for classification.	ISLR 4
10/03	**Classification (II):** support vector machines.	ISLR 9
10/10	NO CLASS: FALL BREAK	
Unsupervised learning I: Clustering		
10/17	**Clustering (I):** k-means; cluster evaluation.	ISLR 10.3
10/24	**Clustering (II):** Gaussian mixture models.	PRML 9.1 – 9.2
Unsupervised learning II: Dimensionality reduction		
10/31	**Dimensionality reduction (I):** principal component analysis.	ISLR 6.3
11/07	**Dimensionality reduction (II):** independent component analysis.	
More statistical learning tools		
11/14	**Decision trees.**	ISLR 8.1
11/21	**Ensemble learning.**	ISLR 8.2
11/28	NO CLASS: THANKSGIVING BREAK	
12/05	**Case studies.**	

ISLR: James, Witten, Hastie, Tibshirani. An Introduction to Statistical Learning: with Applications in R. Springer, 2013.
ESL: Hastie, Tibshirani, Friedman. The Element of Statistical Learning. Springer, 2009.
UDL: Prince. Understanding Deep Learning. MIT Press, 2024.
PRML: Bishop. Pattern Recognition and Machine Learning. Springer New York, 2006.

Pre-course survey: https://pitt.co1.qualtrics.com/jfe/form/SV_7WLqkyvu7u7sWp0.

11. Disability Services Statement

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, drsrecep@pitt.edu, (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

12. Academic Integrity Statement

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on

Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

To learn more about Academic Integrity, visit the [Academic Integrity Guide](#) for an overview of the topic. For hands- on practice, complete the [Understanding and Avoiding Plagiarism tutorial](#).

13. Sexual Misconduct, Required Reporting, and Title IX Statement

If you are experiencing sexual assault, sexual harassment, domestic violence, and stalking, please report it to me and I will connect you to university resources to support you.

University faculty and staff members are required to report all instances of sexual misconduct, including harassment and sexual violence to the Office of Civil Rights and Title IX. When a report is made, individuals can expect to be contacted by the Title IX Office with information about support resources and options related to safety, accommodations, process, and policy. I encourage you to use the services and resources that may be most helpful to you.

As your professor, I am required to report any incidents of sexual misconduct that are directly reported to me. You can also report directly to Office of Civil Rights and Title IX: 412-648-7860 (M-F; 8:30am-5:00pm) or via the Pitt Concern Connection at: [Make A Report](#)

An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as a relevant part of an academic project, classroom discussion, or course assignment, are not required to be disclosed to the University's Title IX office.

If you wish to make a confidential report, Pitt encourages you to reach out to these resources:

- The University Counseling Center: 412-648-7930 (8:30 A.M. TO 5 P.M. M-F) and 412-648-7856 (AFTER BUSINESS HOURS)
- Pittsburgh Action Against Rape (community resource): 1-866-363-7273 (24/7)

If you have an immediate safety concern, please contact the University of Pittsburgh Police, 412-624-2121

Any form of sexual harassment or violence will not be excused or tolerated at the University of Pittsburgh.

For additional information, please visit the [full syllabus statement](#) on the Office of Diversity, Equity, and Inclusion webpage.