Foundations of Machine Learning

MET CS 555 | Section B1-IND (3319) | On-ground (In-Person) | Summer Session Two, 2026
1010 Commonwealth Avenue (MET Building) MET 122

Instructor: Hong Pan, Ph.D.
Email: hongpan@bu.edu
Mobile: (917) 439-2996
Office hours: M/W 5:30–6:00, 9:30–10:00 PM @ MET122 | Weekday 12 - 1 PM 15-Min @ Zoom

Course Description

Learn the foundations of machine learning, regression, and classification by exploring real-world examples. Topics include how to describe data, statistical inference, 1 and 2-sample tests of means and proportions, simple linear regression, multiple linear regression, multinomial regression, logistic regression, analysis of variance, and regression diagnostics. These topics are explored using the statistical package R, with a focus on understanding how to use these methods, interpret their outputs, and visualize the results. In each topic area, the methodology, including underlying assumptions and the mechanics of how it all works, along with appropriate interpretation of the results, is discussed. Concepts are presented in the context of real-world examples to help you learn when and how to deploy different methods.

Building on this foundation, the course extends into modern machine learning methods, including empirical risk minimization (ERM), gradient descent, K-means clustering, naive Bayes classification, the bias-variance tradeoff, and the complete ML pipeline in R using the caret package. Students will implement and compare decision trees, random forests, KNN, Ridge and Lasso regularization, PCA, LDA, SVM, and neural networks. Data visualization and data storytelling are woven throughout.

Prerequisites: MET CS 544, MET CS 550, or the consent of the instructor.

Learning Outcomes

1. Formulate and evaluate data-driven problem-solving approaches by selecting, implementing, and interpreting appropriate supervised and unsupervised machine learning techniques within real-world data contexts.
2. Design and assess end-to-end machine learning workflows by integrating data preparation, model development, validation, and performance evaluation using best practices in data science engineering.

3. Summarize and present data meaningfully, selecting the appropriate analysis for the research questions at hand.
4. Form testable hypotheses, evaluate using statistical analyses, and understand and verify the underlying assumptions of a particular analysis.
5. Conduct, present, and interpret data analyses using R, and communicate analytical insights through visualizations, written reports, and presentations.

Books

Textbooks:

- **An Introduction to Statistical Learning with Applications in R, 2nd ed. (2021)** by James, Witten, Hastie, Tibshirani. Free at https://statlearning.com. **[ISLR]**
- **Introductory Statistics 2e** at OpenStax. **[IS]**
- **Learning Statistics with R** (https://learningstatisticswithr.com/). **[LSR]**
- Customized courseware provided by your instructor. Supplementary video and online resources used in the readings: Against All Odds: Inside Statistics [A]; Crash Course Statistics [C]; Khan Academy Statistics & Probability [K]; 3Blue1Brown [3B1B].

Reference Books:

- Roger Peng. **R Programming for Data Science.** LeanPub, 2014–2022. **[RPDS]**
- Roger D. Peng and Elizabeth Matsui. **The Art of Data Science.** LeanPub, 2015–2018.
- Hadley Wickham et al. **R for Data Science (2e)**. **[Refer to 1e as R4DS1e]**
- Hadley Wickham et al. **ggplot2: Elegant Graphics for Data Analysis (3e)**.
- Max Kuhn and Julia Silge. **Tidy Modeling with R.** https://www.tmwr.org

Courseware

Course website: https://learn.bu.edu (Blackboard)

R and RStudio (or Jupyter Notebook) — install on your laptop before the first class meeting.

Course R code examples on GitHub: https://github.com/kiat/R-Examples

Utilizing the interactive learning tool **swirl** at https://swirlstats.com/ (~15 minutes daily) is recommended during the first few weeks.

Class Policies

1. **Attendance & Absences** — This is an on-campus class. Attendance is mandatory and will be recorded to recognize your commitment to the class. Inform the instructor of any absences in advance to maintain mutual respect.
2. **Assignment Completion & Late Work** — All assignments must be submitted electronically on Blackboard by their due dates. Late submissions are not accepted.
3. **Laptop Requirement** — Students must have a personal laptop. We will use laptops in the classroom to write R programs. Please ensure your computer is fully charged before coming to class.
4. **During class,** cell phones may not be used. Laptops must be put away except during designated activity periods. Tablets may be used for note-taking only if they are flat on the desk. Pencil-and-paper note-taking is encouraged.
5. The best way to contact the instructor is through email. Your questions are valued and will be addressed with care, fostering a supportive learning environment.

   Email subject line: All emails must begin with **[METCS555]** in the subject line. Please use standard professional email etiquette. Email questions should be sent a reasonable time before due dates.

6. **Academic Conduct Code** — Cheating and plagiarism will not be tolerated in any Metropolitan College course. They will result in no credit for the assignment or examination and may lead to disciplinary actions. Please review the Student Academic Conduct Code: https://www.bu.edu/met/current-students/academic-policies-procedures/.
7. **Cooperative Learning** — Group discussion is encouraged, and final grades will not be curved, to promote peer learning. However, write your answers independently. You must understand how to solve the homework problems on your own.

Grading Criteria

Component	Weight	Due Date
Formative	**50%**	**Various**
11 In-Class Handouts + Activities	33% (3% per M)	@End of class meeting
Homework 1	4%	Mon 07/06
Homework 2	4%	Mon 07/13
Homework 3	4%	Mon 07/20
Homework 4	4%	Mon 07/27
Midterm Exam I Reflection	1%	Wed 07/22

Summative	**50%**	**Various**
8 Reading Quizzes	5%	Weekly @Beginning of class meeting, except for M1, 6, 10, 12
Midterm Exam I	10%	Wed 07/15
Midterm Exam II	10%	Wed 07/29
Course Project (Total)	25%	Various
• Proposal	(2.5%)	Wed 07/22
• Progress Report	(2.5%)	Wed 07/29
• Final Report	(15%)	Mon 08/03
• Poster Presentation	(5%)	Wed 08/05
TOTAL	**100%**	

Grading Scale

Boston University uses the honor point system. See https://www.bu.edu/reg/academics/grades-gpa/.

Grade	Honor Points	Percentage	Description
A	4.0	93–100%	Excellent
A–	3.7	90–92.9%	Excellent
B+	3.3	87–89.9%	Excellent
B	3.0	83–86.9%	Good
B–	2.7	80–82.9%	Good
C+	2.3	77–79.9%	Good
C	2.0	73–76.9%	Satisfactory
C–	1.7	70–72.9%	Satisfactory
D	1.0	60–69.9%	Low, pass
F	0.0	< 60%	Fail, no credit

Class Meetings, Lectures & Assignments

Location: 1010 Commonwealth Avenue (MET Building) MET 122 | Mon & Wed, 6:00–9:30 PM ET | June 29 – August 7, 2026

Lectures, Readings, and assignments are subject to change and will be announced in class within a reasonable time frame.

#	Date	Topics	Reading	Deliverables
Module I: Descriptive Statistics, Sampling & Experimental Design, Probability Basics & CLT				
1	Mon 06/29	Introduction & One-Variable Descriptive Statistics	A 1–6; C 1–5; K Units 1–3; IS 1.1–1.3, 2; LSR 5–6; ISLR 1	—
2	Wed 07/01	Two-Variable Descriptive Statistics: Z-Scores, Normal Distribution, Correlation, LSRL	A 7–17; C 5–12; K Units 4–6; IS 1.4–1.6, 12.1–12.3; LSR 2, 5–7; ISLR 2.1	—
3	Mon 07/06	Sampling & Experiment Design | Probability, Random Variables, Sampling Distributions, CLT	A 15–24; C 9–20; K Units 6–11; IS 3–8; LSR 9–10	HW 1 Due
Module II: Statistical Inference, ANOVA, Multiple Regression, GLM, Advanced Methods				
4	Wed 07/08	Statistical Inference for Proportions (CIs, Hypothesis Tests, Chi-Square)	A 24, 25, 28, 29; C 20–23, 26, 28–30; K Units 11–14; IS 8–11; LSR 10–13	—
5	Mon 07/13	Statistical Inference for Means (One-Sample, Two-Sample, Paired) & Regression Slope	A 25–27, 30; C 20–23, 26–28, 32; K Units 11–13, 15; IS 8–10, 12; LSR 10, 11, 13	HW 2 Due
6	Wed 07/15	Midterm I (75 min) | ANOVA, Multiple Regression, GLM	A 31; C 33–35; K Unit 16; IS 13; LSR 14–16; ISLR 3	Midterm I

7	Mon 07/20	Logistic Regression | Nonparametric Methods, Bootstrap & Permutation Tests	ISLR 4.3, 4.6, 5.2	HW 3 Due
Module III: Machine Learning Foundations, Methods, Pipelines, Workflow, AI Preview				
8	Wed 07/22	ML Foundations: Loss Functions, ERM, Gradient Descent, K-Means, Naive Bayes, Bias-Variance Tradeoff	ISLR 2.1–2.2, 10.3, 4.4.1; R4DS 1–3; K (K-means & Naïve Bayes tutorials); 3B1B (Gradient descent)	Proposal Due
9	Mon 07/27	ML Pipeline (caret): Decision Trees, Random Forests, KNN, Ridge & Lasso, Data Viz & Storytelling	ISLR 6.2, 8.1–8.2; R4DS 2	HW 4 Due
10	Wed 07/29	Midterm II (75 min) | PCA, LDA, BN, SVM	ISLR 9.1–9.3, 12.1–12.2, 4.4; 3B1B (But what is a neural network?)	Midterm II; Progress Rpt
11	Mon 08/03	Neural Networks & Applications | Course Review	ISLR 10.1–10.3	Final Report Due
12	Wed 08/05	Final Project Poster Presentations & Peer Evaluation | Deep Learning Preview	ISLR 10.6–10.7	Poster Pres.

Tips for Success: Evidence-Based Learning Strategies

Research on learning science shows that intelligence alone is not enough—graduate school requires DIFFERENT learning strategies. These are proven to produce significantly better results than passive review.

What Actually Works:

- **Retrieval Practice:** Test yourself BEFORE looking at answers. Close notes, try to recall, do problems without help first, and explain out loud.
- **Distributed Practice:** Spread study over days, not hours. 1 hr/day × 5 days > 5 hrs in one day. Cramming = weak memory.

- **Interleaving:** Mix different topics in each study session rather than blocking one topic at a time.
- **Elaborative Rehearsal:** Connect new info to what you know: “This is like...” “This relates to...”
- **Self-Testing:** Quiz yourself constantly. Cover the answers and try to solve. Make your own flashcards. Explain to a friend.
- **Metacognition:** “Do I REALLY understand this, or just recognize it?” “Can I explain this to someone?”

What Does NOT Work: Re-reading notes, highlighting everything, watching without doing, copying solutions, marathon sessions. **Key Insight:** Struggle = Learning. If it feels easy, you’re probably not learning much.

The PLAN-WORK-REFLECT Cycle:

- **Before:** Set goals
- **During:** Engage actively (do problems, explain, test yourself)
- **After:** Reflect and adjust

Time Commitment: 12–15 hrs/week outside class. Daily practice (15+ min) makes a big difference. **FACE:** Focus–Attitude–Creativity–Effort. **Growth Mindset:** Your brain grows with exercise! Feeling confused is GOOD.

“The secret of getting ahead is getting started.” — Mark Twain.

Statement of Support

Take care of yourself. Maintain a healthy lifestyle this summer. If you or anyone you know experiences academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support.

Our Classroom Community

At Boston University, faculty and students collaborate to create a respectful and inclusive learning environment where the diversity of backgrounds and perspectives of all members is valued and respected.

Diversity and Inclusion

Boston University is committed to equal opportunity, affirmative action, diversity, and social justice. In the classroom, we cultivate an inclusive environment that denounces discrimination through innovation, collaboration, and an awareness of global perspectives on social justice.

Syllabus Statement

This syllabus is not a contract. The instructor reserves the right to modify course requirements and/or assignments in response to new materials, class discussions, or other legitimate pedagogical objectives.