Boston University Metropolitan College

Foundations of Machine Learning

MET CS555 A1

Guanglan Zhang
guanglan@bu.edu
Office hours: Thursday 10 - 11:30 AM or by appointment in Room 329, 1010 Commonwealth
Class Time & Location: 12:30 - 3:15 PM Thursdays, Room 217, 640 Commonwealth Avenue

Course Description

The content of this course is twofold. One aspect is a theoretical aspect, in which the course provides foundations of statistical machine learning and covers fundamental basic blocks of data analytics, hypothesis testing, regression, and classification. The other aspects are learning the statistical package R, coding in R, implementing theories on real and semi-real data, and examining theories learned in the course in action.

The course topics include describing data, statistical inference, 1 and 2 sample tests of means and proportions, simple linear regression, multiple 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 and interpret output from this software as well as how to visualize results. In each topic area, the methodology, including underlying assumptions and the mechanics of how it works, along with the appropriate interpretation of the results, are discussed. Concepts are presented in the context of real-world examples.

Prerequisites

CS544 Foundations of Analytics and Data Visualization or or CS550 Computational Mathematics for Machine Learning. Or equivalent background.

Learning Objectives

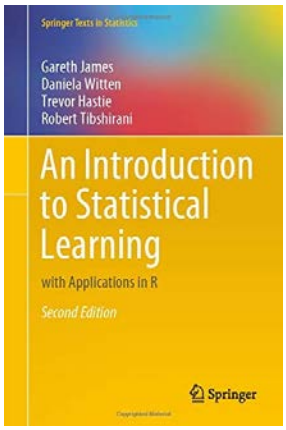
By successfully completing this course, you will be able to do the following:

- Understand the basics of regression and classification algorithms.
- Assess a regression or a classification algorithm.
- Define many problems in an objective manner and in such a way that the results can be evaluated
- Select the appropriate statistical analysis depending on the question at hand.
- Describe/Verify the underlying assumptions of a particular analysis.
- Data manipulation using R.
- Summarize and present data in meaningful ways.
- Conduct, present, and interpret common statistical analyses using R.
- Communicate results to others effectively and clearly.

Boston University Metropolitan College

Recommended Book

These books should be used as a reference to help support you in your learning and supplement the classroom sessions.

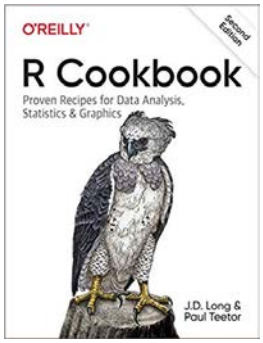


James, G., Witten, D., Hastie, T,. Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics), 2nd edition.

Publisher: Springer

ISBN: 978-1071614174

It is freely available online at https://www.statlearning.com/

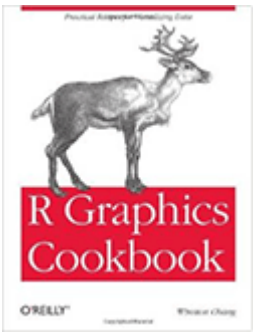


Long, J. D. & Teetor, P. (2019). R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics, 2nd edition.

Publisher: O'Reilly Media

ISBN: 978-1492040682.

It is freely available online at: https://rc2e.com/.



Chang, W. (2021). An Introduction to Statistical Learning: with Applications in R, 2nd edition.

Publisher: O'Reilly Media

ISBN: 978-1491978603

It is freely available online at: https://r-graphics.org/

Courseware

There are six online modules covering the course content in the Blackboard site.

Grading Criteria

- Homework Assignments and Term Project

The six homework assignments are focused on applying theory learned in the week's module to a set of data and analyzing that data in R. Assignment submissions should be a single Microsoft Word or PDF file. The R code used to generate your results should be appended to the end of your assignment. Term project at the end of the semester gives you freedom to select a research question of your interest and answer it by applying what you have learnt in the course.

- Quizzes
  The quizzes will evaluate students understanding of concepts presented in the corresponding week's module. Students should ensure adequate preparation before starting the quiz. It will not be possible to do well on the quiz without first reviewing the course material in depth and attempting to understand all examples and test yourself questions. It is recommended that you complete the quiz after you feel comfortable with the material and asked any questions that you may have had.

- Midterm Examination and Final Examination
  The midterm exam will cover material from module 1-3. The final exam will be comprehensive and will cover material from the entire course. Both are close-book and close- notes exams.

The final grade for this course will be based on the following:

Deliverable	**Weight**
Assignments	25%
Quizzes	15%
Midterm Exam	20%
Class Participation	5%
Term Project	10%
Final Exam	25%

Class Meetings, Lectures & Assignments

Lectures, Readings, and Assignments subject to change, and will be announced in class as applicable within a reasonable time frame.

Date	Topic	Readings Due	Due dates
Lecture 1 1/22	• Fundamental Elements of Statistics • Qualitative and Quantitative Data Summaries	Online Module 1 Class slides	
Lecture 2 1/29	• Normal distribution • Sampling • The Central Limit Theorem	Online Module 1 Class slides	

Lecture 3 2/5	• Statistical Inference • Confidence Intervals • Test of Significance • Stating Hypotheses • Test Statistics and p-Values • Evaluating Hypotheses	Online Module 2 Class slides	Assignment 1 Quiz 1
Lecture 4 2/12	• Significance Test "Recipe" • Significance Tests and Confidence Intervals • Inference about a Population Mean • Two-Sample Problems	Online Module 2 Class slides	
Lecture 5 2/19	• Scatterplots • Correlation	Online Module 3 Class slides	Assignment 2 Quiz 2
Lecture 6 2/26	• Simple Linear Regression • F-test for Simple Linear Regression • t-test for Simple Linear Regression	Online Module 3 Class slides	
Lecture 7 3/5	• Residual Plots • Outliers and Influence Points • Assumptions of least-square regression	Online Module 4 Class slides	Assignment 3 Quiz 3
Midterm 3/19	Midterm Exam		
Lecture 8 3/26	• Equation of multiple linear regression • Interpretation of multiple linear regression • F-test for Multiple Linear Regression • t-tests in Multiple Linear Regression • Cautions about Regression	Online Module 4 Class slides	
Lecture 9 4/2	• One-Way Analysis of Variance • F-test for ANOVA • Evaluating Group Differences • Type I and Type II Errors	Online Module 5 Class slides	Assignment 4 Quiz 4
Lecture 10 4/9	• Issues with Multiple Comparisons • Assumptions of Analysis of Variance • Relationship between One-Way Analysis of Variance and Regression • One-Way Analysis of Covariance • Two-Way Analysis of Variance	Online Module 5 Class slides	

	• Two-Way Analysis of Covariance		
Lecture 11 4/16	• One-Sample Tests for Proportions • Significance Tests for a Proportion • Confidence Intervals for a Proportion	Online Module 6 Class slides	Assignment 5 Quiz 5
Lecture 12 4/23	• Two-Sample Tests for Proportions • Confidence Intervals for Differences in Proportions • Significance Tests for Differences in Proportions • Effect Measures • Logistic Regression • Multiple Logistic Regression • Area under the ROC Curve	Online Module 6 Class slides	
Lecture 13 4/30	Project presentations		Assignment 6 Quiz 6 Term Project
5/7	Final Exam		

Instructor Biography
Guanglan Zhang, Ph.D.

Dr. Guanglan Zhang received her Ph.D. from the School of Computer Engineering, Nanyang Technological University, Singapore, for doctoral work in bioinformatics. She is an associate professor of computer science department at Boston University Metropolitan College. Dr. Zhang has worked in the data mining and data analytics field since 1998. The most important aspects of her work include biomedical data analysis, development, and implementation of biomedical databases, computational simulations of laboratory experiments, development of diagnostic methods for tissue typing, and computational support for vaccine development. Computational tools that she developed are used in the study of immunology, vaccinology, infectious disease, and cancer. She has authored more than 50 peer-reviewed scientific journal publications and developed dozens of biomedical and computational systems.

CLASS POLICIES

1) **Attendance & Absences** –

Students should plan to attend all classes in-person (see "Class Meetings, Lectures & Assignments"). If an absence is necessary, please plan with course professor in advance.

2) Assignment Completion & Late Work

All quizzes and assignments have to be submitted by the due dates. Each 24 hours of delay will result in 10% penalty unless rescheduling has been permitted by course instructor. Class projects need to be completed and presented by the due date.

3) Academic Conduct Code

For the full text of the academic conduct code, please go to http://www.bu.edu/met/for-students/met-policies-procedures-resources/academic-conduct-code/.

A Definition of Plagiarism

"The academic counterpart of the bank embezzler and of the manufacturer who mislabels products is the plagiarist: the student or scholar who leads readers to believe that what they are reading is the original work of the writer when it is not. If it could be assumed that the distinction between plagiarism and honest use of sources is perfectly clear in everyone's mind, there would be no need for the explanation that follows; merely the warning with which this definition concludes would be enough. But it is apparent that sometimes people of goodwill draw the suspicion of guilt upon themselves (and, indeed, are guilty) simply because they are not aware of the illegitimacy of certain kinds of "borrowing" and of the procedures for correct identification of materials other than those gained through independent research and reflection."

"The spectrum is a wide one. At one end there is a word-for-word copying of another's writing without enclosing the copied passage in quotation marks and identifying it in a footnote, both of which are necessary. (This includes, of course, the copying of all or any part of another student's paper.) It hardly seems possible that anyone of college age or more could do that without clear intent to deceive. At the other end there is the almost casual slipping in of a particularly apt term which one has come across in reading and which so aptly expresses one's opinion that one is tempted to make it personal property."

"Between these poles there are degrees and degrees, but they may be roughly placed in two groups. Close to outright and blatant deceit-but more the result, perhaps, of laziness than of bad intent-is the patching together of random jottings made in the course of reading, generally without careful identification of their source, and then woven into the text, so that the result is a mosaic of other people's ideas and words, the writer's sole contribution being the cement to hold the pieces together. Indicative of more effort and, for that reason, somewhat closer to honest, though still dishonest, is the paraphrase, and abbreviated (and often skillfully prepared) restatement of someone else's analysis or conclusion, without acknowledgment that another person's text has been the basis for the recapitulation."

The paragraphs above are from H. Martin and R. Ohmann, The Logic and Rhetoric of Exposition, Revised Edition. Copyright 1963, Holt, Rinehart and Winston.

Academic Conduct Code

I. Philosophy of Discipline

The objective of Boston University in enforcing academic rules is to promote a community atmosphere in which learning can best take place. Such an atmosphere can be maintained only so long as every student believes that his or her academic competence is being judged fairly and that he or she will not be put at a disadvantage because of someone else's dishonesty. Penalties should be carefully determined so as to be no more and no less than required to maintain the desired atmosphere. In defining violations of this code, the intent is to protect the integrity of the educational process.

II. Academic Misconduct

Academic misconduct is conduct by which a student misrepresents his or her academic accomplishments, or impedes other students' opportunities of being judged fairly for their academic work. Knowingly allowing others to represent your work as their own is as serious an offense as submitting another's work as your own.

III. Violations of this Code

Violations of this code comprise attempts to be dishonest or deceptive in the performance of academic work in or out of the classroom, alterations of academic records, alterations of official data on paper or electronic resumes, or unauthorized collaboration with another student or students. Violations include, but are not limited to:

A. Cheating on examination. Any attempt by a student to alter his or her performance on an examination in violation of that examination's stated or commonly understood ground rules.
B. Plagiarism. Representing the work of another as one's own. Plagiarism includes but is not limited to the following: copying the answers of another student on an examination, copying or restating the work or ideas of another person or persons in any oral or written work (printed or electronic) without citing the appropriate source, and collaborating with someone else in an academic endeavor without acknowledging his or her contribution. Plagiarism can consist of acts of commission-appropriating the words or ideas of another-or omission failing to acknowledge/document/credit the source or creator of words or ideas (see below for a detailed definition of plagiarism). It also includes colluding with someone else in an academic endeavor without acknowledging his or her contribution, using audio or video footage that comes from another source (including work done by another student) without permission and acknowledgement of that source.
C. Misrepresentation or falsification of data presented for surveys, experiments, reports, etc., which includes but is not limited to: citing authors that do not exist; citing interviews that never took place, or field work that was not completed.
D. Theft of an examination. Stealing or otherwise discovering and/or making known to others the contents of an examination that has not yet been administered.
E. Unauthorized communication during examinations. Any unauthorized communication may be considered prima facie evidence of cheating.
F. Knowingly allowing another student to represent your work as his or her own. This includes providing a copy of your paper or laboratory report to another student without the explicit permission of the instructor(s).

G. Forgery, alteration, or knowing misuse of graded examinations, quizzes, grade lists, or official records of documents, including but not limited to transcripts from any institution, letters of recommendation, degree certificates, examinations, quizzes, or other work after submission.
H. Theft or destruction of examinations or papers after submission.
I. Submitting the same work in more than one course without the consent of instructors.
J. Altering or destroying another student's work or records, altering records of any kind, removing materials from libraries or offices without consent, or in any way interfering with the work of others so as to impede their academic performance.
K. Violation of the rules governing teamwork. Unless the instructor of a course otherwise specifically provides instructions to the contrary, the following rules apply to teamwork: 1. No team member shall intentionally restrict or inhibit another team member's access to team meetings, team work-in-progress, or other team activities without the express authorization of the instructor. 2. All team members shall be held responsible for the content of all teamwork submitted for evaluation as if each team member had individually submitted the entire work product of their team as their own work.
L. Failure to sit in a specifically assigned seat during examinations.
M. Conduct in a professional field assignment that violates the policies and regulations of the host school or agency.
N. Conduct in violation of public law occurring outside the University that directly affects the academic and professional status of the student, after civil authorities have imposed sanctions.
O. Attempting improperly to influence the award of any credit, grade, or honor.
P. Intentionally making false statements to the Academic Conduct Committee or intentionally presenting false information to the Committee.
Q. Failure to comply with the sanctions imposed under the authority of this code.

Disability Services

In accordance with University policy, every effort will be made to accommodate unique and special needs of students with respect to speech, hearing, vision, or other disabilities. Any student who feels he or she may need an accommodation for a documented disability should contact the Office of Disability Services (http://www.bu.edu/disability) at (617) 353-3658 or at access@bu.edu for review and approval of accommodation requests.

Etiquette and Netiquette

Before posting to any discussion forum, sending email, or participating in any course or public area, please consider the following:

Ask Yourself…

- How would I say this in a face-to-face classroom or if writing for a newspaper, public blog, or wiki?
- How would I feel if I were the reader?
- How might my comment impact others?
- Am I being respectful?
- Is this the appropriate area or forum to post what I have to say?

When you are speaking or writing, please follow these rules:

- Stay polite and positive in your communications. You can and should disagree and participate in discussions with vigor; however, when able, be constructive with your comments.
- Proofread your comments before you post them. Remember that your comments are permanent.
- Pay attention to your tone. Without the benefit of facial expressions and body language your intended tone or the meaning of the message can be misconstrued.
- Be thoughtful and remember that classmates' experience levels may vary. You may want to include background information that is not obvious to all readers.
- Stay on message. When adding to existing messages, try to maintain the theme of the comments previously posted. If you want to change the topic, simply start another thread rather than disrupt the current conversation.
- When appropriate, cite sources. When referencing the work or opinions of others, make sure to use correct citations.

When you are reading your peers' communication, consider the following:

- Respect people's privacy. Don't assume that information shared with you is public; your peers may not want personal information shared. Please check with them before sharing their information.
- Be forgiving of other students' and instructors mistakes. There are many reasons for typos and misinterpretations. Be gracious and forgive other's mistakes or privately point them out politely.
- If a comment upsets or offends you, reread it and/or take some time before responding.

Important Note

Don't hesitate to let your instructor or student services coordinator know if you feel others are inappropriately commenting in any forum.

All Boston University students are required to follow academic and behavioral conduct codes. Failure to comply with these conduct codes may result in disciplinary action.