

Syllabus for Mathematics of Data Science

AS.110.205

Instructor Information

Professor Ratigan

Email: cratiga1@jhu.edu this is by far the best way to contact me.

Zoom: <https://jhu-cty.zoom.us/my/cratigan> Passcode: 918057

Office Hours: To be determined at the start of semester, otherwise by appointment

Course Materials

The course will require internet access for Zoom office hours and accessing the course website. Additionally, the course will require the student to work with python notebooks through Jupyter Notebooks using a miniconda distribution and requires the textbook: *An Introduction to Statistical Learning with applications in Python* (available for [free online](#)).

Course Schedule

The course is thirteen weeks broken up into the following content, note the readings will skip around a bit in the textbook.

Week	Topics Covered.
Week 1	Introduction to programming: Command Line, Github+Git and Python basics: datatypes, list comprehension, slicing, control flow, recursion, etc.
Week 2	Overview of Machine Learning: Supervised/Unsupervised Learning, Classification, Regression. Pandas, Matplotlib and Numpy
Week 3	Model Evaluation: Performance Metrics, Probability, Hypothesis Testing.
Week 4	Advanced Regression Models; train_test_split/cross_validation; Project 1: Modeling Housing Prices
Week 5	Unsupervised Learning Networks, Clustering + Performance Metrics.
Week 6	Principal Component Analysis, Dimensionality Reduction, Spectral Clustering
Week 7	Neural Networks: Perceptrons, Newton's Method, Backpropagation.
Week 8	Convolutional Neural Networks, machines that learn, project 2: image processing.
Week 9	Advanced Modeling Regimes: Pipelines and gridsearching
Week 10	Text, Timeseries and Sound data, the Fourier Transform

Week 11	Recurrent Neural Networks,
Week 12	Recommender Systems
Week 13	Bias-Variance Tradeoff, additional topics, final project week.

Online Lectures

Prerecorded lectures will be posted for the week in each module together with interactive python notebook (.ipynb) files for you to follow along in. Installation instructions for using .ipynb files can be found in the course information module and will be handled during the orientation session.

There will be live synchronous sessions throughout the semester through zoom, with dates posted in the course. During these live synchronous components, homework problems and project review will take place. Links to the recordings of each live online sessions will be posted in Canvas and emailed to the class.

For more information regarding Zoom, please see the Zoom Student Quick Start Guide.

Course Description

This course is designed for students of all backgrounds to provide a solid foundation in the underlying mathematical, programming, and statistical theory of data science. In today's data driven world, it is increasingly important to be data literate. To this end, the course will motivate the fundamental concepts used in this growing field. While discussing the general theory behind common methods of data analysis, there will be numerous applications to real world data sets. In particular, the course will use Python libraries to create, import, and analyze data sets. There are no mathematical prerequisites for this course, though prior knowledge of calculus, statistics and/or programming can be helpful.

Course Goals

This course will prepare you to understand the opportunities, challenges and limitations of data science and machine learning. By teaching you mastery in the techniques, theory, and applications of the data science process this course will serve as a stepping stone to navigating real world issues in our AI driven world.

Course Objectives

By the conclusion of this course, you are expected to have gained the ability to

- Understand the theory underlying modern data science
- Fit and train a model to solve regression and classification problems
- Use various metrics to evaluate the effectiveness of a model
- Understand the distinction between regression and classification, supervised and unsupervised learning.
- Communicate effectively regarding real world data.

Course Textbooks

Required: *An Introduction to Statistical Learning* by Gareth James et al. (available for [free online](#))

Optional: *The Grammar of Graphics* by Leland Wilkinson ISBN 0387245448 (recommended)

Student Coursework Requirements

Homework

Will be assigned each week as Python Notebook problem sets, these are to be completed and then uploaded to Canvas for review.

Projects

There will be three projects applying the skills students developed to actual datasets. The first project will have students analyze a well-known dataset, the final project will have the students work on a modeling problem of their choice.

Quizzes

There will be a short (20 minute) online quiz each week. There are two attempts at the quiz and the questions are pulled randomly during each attempt. The higher of the two attempts counts for the grade. Your lowest quiz grade will be dropped.

Discussion Forums

There will be weekly discussion forums to give students the opportunity to read and write about the material in the course.

Grading

The final grade for the class will be given as a weighted average as follows.

Homeworks: 25% (lowest dropped)

Project 1: 15%

Project 2: 15%

Project 3: 15%

Quizzes: 20% (lowest dropped)

Discussion Forums: 10%

The letter grades are assigned as follows based on your final weighted average:

A: 90-100

B: 80 - 89

C: 70 - 79

D: 63 - 69

F: < 63

"+" and "-" will be determined at the end of the semester.

Students with disabilities

Students with documented disabilities or other special needs who require accommodation must register with Student Disability Services. After that, remind the instructor of the specific needs at least one week prior to each exam; the instructor must be provided with the official letter stating all the needs from Student Disability Services.

There may be a student in this class who requires the services of a note taker. This is an opportunity to share notes through the Student Disability Services Office. If you are interested in performing this service, please register as a notetaker with Student Disability Services.

Academic Integrity

Academic Misconduct Policy

All students are required to read, know, and comply with the Johns Hopkins University Krieger School of Arts and Sciences (KSAS) / Whiting School of Engineering (WSE) Procedures for Handling Allegations of Misconduct by Full-Time and Part-Time Graduate Students.

This policy prohibits academic misconduct, including but not limited to the following: cheating or facilitating cheating; plagiarism; reuse of assignments; unauthorized collaboration; alteration of graded assignments; and unfair competition. You may request a paper copy of this policy at this by contacting jhep@jhu.edu.

JHU ethics statement

"Undergraduate students enrolled in the Krieger School of Arts and Sciences or the Whiting School of Engineering at the Johns Hopkins University assume a duty to conduct themselves in a manner appropriate to the University's mission as an institution of higher learning. Students are obliged to refrain from acts which they know, or under circumstances have reason to know, violate the academic integrity of the University. [The JHU Code of Ethics]"

Ethical violations include cheating on exams, plagiarism, reuse of assignments, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition.