Data Science for Physics

Edit

Course number and section: **Phys 4803-A / Phys 6262-A**
Semester: **Fall 2026**
Classroom: **Howey S-106**
Classes: **Mo/We 2:00 - 3:15 PM**
Registrar assigned Final exam day: December 16, 2:40 - 5:30 PM

Instructor

Ignacio Taboada
Office: **N-116 Howey Building**
Email: **itaboada@gatech.edu (mailto:itaboada@gatech.edu)**
Office Hour: Wednesday 3:15 - 4:15 PM

Course description

This class is dedicated to Data Science. The main three areas to be studied are: Probability, Statistics and Machine Learning. The class is joint for graduate and undergraduate students, with differences in expectations, objectives and evaluation.

For physics/astrophysics majors: this course can count as a technical elective.

A minimal knowledge (variables and data types, flow control) of Python is expected. Knowledge of multivariate calculus and linear algebra is assumed. Some course activities assume familiarity with physics, such as quantum mechanics or Newtonian physics.

Course objectives and learning outcomes

Understand the definitions of probability. Be able to apply Bayes theorem to physical processes. Develop skills to propose Bayesian priors (Grad students).

Identify when a physical process is described by a given Integer Probability distribution. Identify when a random process is described by a probability density in one or multiple dimensions.

Be able to use maximum likelihood and chi-squared in regression. Be able to use likelihood ratio in binomial hypothesis testing. Be able to propose a Likelihood function for a given physical process (Grad students).

Acquire skills in designing decision trees. Develop skills to identify over-training in machine learning. Design and train random forests and boosted decision trees (graduate students).

Be able to design and implement a Multi-layer Perceptron neural network (plain vanilla neural network). Diagnose over-training visually with loss curves. Diagnose over-training quantitatively (graduate students). Acquire skills to propose a physics-informed loss function (graduate students).

Textbook

Statistical Methods for Data Analysis in Particle Physics.
Author: Luca Lista
3rd Edition. (There ARE differences with respect to prior editions).
ISBN 978-3-031-19934-9
Available for download at library.gatech.edu
No particle physics knowledge is required for this course.

Additional resources

Probability in Physics, an introductory guide
Andy Lawrence
ISBN-13 978-3030045425

Statistical Data Analysis
Glen Cowan
ISBN-13 978-0-198-50155-8

Statistics for nuclear and particle physicists
Louis Lyons
ISBN-13 978-1316100097

YouTube videos by 3Blue1Brown:
https://www.3blue1brown.com/ (https://www.3blue1brown.com/)

Notes on Contemporary Machine Learning for Physicists
Jared Kaplan:
https://smaeland.github.io/reference-manager/pdfs/ContemporaryMLforPhysicists.pdf (https://smaeland.github.io/reference-manager/pdfs/ContemporaryMLforPhysicists.pdf)

Grading Policy

- Quizzes 5%
- In-class activities: 15%
- Project 1: 15% - Probability density functions
  Due Wednesday October 7. 10:00 PM
- Project 2: 15% - Regression
  Due Wednesday October 21. 10:00 PM

- Project 3: 15% - Hypothesis testing
  Due Wednesday November 4. 10:00 PM
- Project 4: 15% - Classification with Decision Trees and Neural Networks
  Due Tuesday November 22, 10:00 PM
- Oral Examination: 20% - Undergraduates, see class schedule. Graduate students: day assigned for final exam by the Registrar.

Grading Scale: 90 - 100 = A; 80 - 89 = B; 70 - 79 = C; 60 - 69 = D; 0 - 59 = F.

A "D" or better is passing for those taking the course pass/fail. For people auditing the class, 85% participation is required for "V".

Grades are not "curved".

Quizzes

Every flipped class will begin with a quiz. The quiz will be administered using the Point Solutions app. You should install this app in your phone (android or iOS). You should link your Georgia Tech email with Point Solutions by logging into **PointSolutions (https://account.turningtechnologies.com/account/authenticate/index)** .

In-Class activities

In class activities are the main activity during flipped classes. You should upload your answers via canvas. Answers can be written on sheets of paper or as Jupyter notebooks

Projects

Projects should be submitted as Jupyter notebooks.

Project Review

The project review will test the understanding the student of the projects they have submitted. The Instructor will ask conceptual question on any of the 4 projects. The instructor will ask questions on code implementation on any of the projects. The correctness of the project will not be reviewed, instead the understanding of the project by the student will be reviewed.

If, in the opinion of the instructor, the Project Review indicates that the student does not understand the material submitted, the instructor may assign a longer oral examination during the slot assigned for the Final exam of this course.

Grading Rubric

The table below shows potential grade demerits per activity (quizzes, projects, in class activities, etc.)

Grading Rubric

Percent	If…
-10%	Figures / plots / tables are not properly presented, and/or code is not properly documented, and/or there are significant deficiencies in coding style. A maximum demerit of 10% will be applied per activity (no double jeopardy). See text for details.
-20%	The student understands the main concepts and problem-solving techniques but has some minor yet non-trivial gaps in their reasoning.
-40%	The student has partially understood the problem. The student is not completely lost but requires tutoring in some of the basic concepts. The student may have started out correctly but took an incorrect path or not finished the problem.
-60%	The student has a poor understanding of the problem. The student may have gone in a not-entirely-wrong but unproductive direction or attempted to solve the problem using pattern matching or by rote memorization.
-80%	The student did not understand the problem. They may have written some appropriate formulas or diagrams but nothing further. Or they may have done something entirely wrong.
-100%	The student wrote nothing or almost nothing. Or what is written is entirely irrelevant to the question.

Plots, figures, code documentation, coding style

These are the expectations for plots, figures, code documentation and coding style

- Label every axis on plots.
- Use appropriate units on every quantity on plots.
- If you use a quantity with units on Python code, you must provide a comment stating the unit used.

- Variable names must be understandable.
- When a project asks for a quantity to be calculated, don't simply print the quantity to the screen. Instead, print out a description of the quantity being printed.
- When more than one quantity is plotted a legend must be provided.
- Code must be sufficiently commented to be reviewable/gradable (explain what is that you are doing).
- The use of vectorization with numpy is strongly encouraged. Egregious use of for loops, when numpy enables vectorization can result in a demerit.

Attendance and Participation Policy

Attendance will be recorded and will count towards grade.

Participation and attendance are critical to achieving course objectives. For this, and many other courses, there's a clear difference in performance between students who regularly attend class and those who don't.

Academic integrity statement

Students are expected to act ethically. Review Georgia Tech's Honor Code and the student Code of Conduct. Any student suspected of cheating or plagiarism on a quiz, exam, project or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Student-Faculty Expectation

I expect an atmosphere of mutual respect and responsibility between faculty and students. Students shall be able to ask any topic-appropriate question in class or office hour and receive a judgement free answer by the faculty or other students.

Services offered through the Office of Disability Services

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services (404-894-2563) as soon as possible to make an appointment
to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Schedule

Schedule

August 24 **Syllabus. How to study for this class. Expectations.**	August 26 **Probability (https://gatech.instructure.com/courses/559400/files/folder/Reading%20Material?**

	preview=75648733) : **Kolmogorov's axioms. Conditional Probability. Independence. Law of total probability.** **Flipped**
August 31 **Probability (https://gatech.instructure.com/courses/559400/files/folder/Reading%20Material?preview=75648733) :** **Bayes Theorem. Classical (Symmetry) probability.** **Flipped**	September 2 **Probability (https://gatech.instructure.com/courses/559400/files/folder/Reading%20Material?preview=75648733) :** **Frequentist and Bayesian Probability. Average, variance, etc.** **Flipped**
September 7 **Labor day**	September 9 **Probability Mass Functions (https://gatech.instructure.com/courses/559400/files/folder/Reading%20Material?preview=75649167) :** **Bernoulli and Binomial** **Flipped**
September 14 **Probability Mass Functions (https://gatech.instructure.com/courses/559400/files/folder/Reading%20Material?preview=75649167) :** **Poisson and Normal** **Flipped**	September 16 **Probability Density Functions**
September 21 **CTEL survey** **Probability Density Functions**	September 23 **Multi-Dimension PDFs**
September 28 **Multi-Dimension PDF**	September 30 **Parameter Estimation**
October 5 **Maximum Likelihood** Ch 6	October 7 **Maximum Likelihood** Ch. 6
October 12 **Chi-squared**	October 14 **Chi-squared**

Ch. 6	Ch. 6
October 19 **Fall Break**	October 21 **Hypothesis Testing**
October 26 **Hypothesis testing**	October 28 **Hypothesis testing**
November 2 **Decision Trees**	November 4 **Decision Trees**
November 9 **Decision Trees**	November 11 **Neural Networks**
November 16 **Neural Networks**	November 18 **Neural Networks**
November 23 **Ugrad Project Review**	November 25 **Thanksgiving break**
November 30 **Ugrad Project Review**	December 2 **Ugrad Project Review**
December 7 **Course Review**	

Course Summary:

Date	Details	Due