Data Science for Physicists

Edit

Course number and section: Phys 4803-A / Phys-6262-IT
Semester: Fall 2026
Classroom: TBD

Instructor

Instructor: Ignacio Taboada
Office: N-116 Howey Building
Email: itaboada@gatech.edu (mailto:itaboada@gatech.edu)
Office Hour: TBD.

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.

The course will be conducted approximately 50% as lectures and 50% in a flipped teaching model.

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 (https://www.amazon.com/Statistical-Methods-Analysis-Particle-Physics/dp/3319628399/ref=sr_1_3?dchild=1&keywords=Statistical+Methods+for+Data+Analysis+in+Particle+Physics&qid=1621954722&sr=8-3) 978-3-031-19934-9
Available for download at library.gatech.edu
No particle physics knowledge is required for this course.

Additional resources

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

Statistics for nuclear and particle physicists
Louis Lyons
ISBN 0-521-37934-2

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

Grading Policy

For undergraduate students

- Quizzes: 10% of the course grade
- Participation: 10% of the course grade
- Class activities: 80%

For graduate students

- Quizzes: 10% of the course grade
- Participation: 10% of the course grade
- Class activities: 60%
- Final project: 20%

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".

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.

On each "flipped" class day, a student will be randomly selected and they will dedicate 10 minutes to describe their results.

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

August 24 Syllabus. How to study for this class. Expectations	August 26 Flipped A minimal introduction to Python.
August 31 Flipped A minimal introduction to numpy	September 2 **Probability.** Kolmogorov's axioms. Conditional Probability. Independent random events.

	Ch. 1 and Ch. 5
September 7 **Labor day**	September 9 **Probability.** Law of total probability. Bayes Theorem. Symmetry and probability. Ch 1 and Ch. 5
September 14 Flipped **Probability.** Frequentist probability. Bayesian probability. Ch.1 and additional material	September 16 Flipped Probability Distributions Ch.2
September 21 Probability Density Functions Ch. 3	September 23 Flipped Probability Density Functions Ch 3
September 28 Probability Density Functions in multiple dimensions Ch 3	September 30 Flipped Probability Density Functions in multiple dimensions Ch 3
October 5 Parameter Estimation Ch 6	October 7 Flipped Maximum Likelihood Ch. 6
October 12 Maximum Likelihood Ch. 6	October 14 Chi-squared Ch. 6
October 19 **Fall Break**	October 21 Flipped Chi-squared Ch. 6