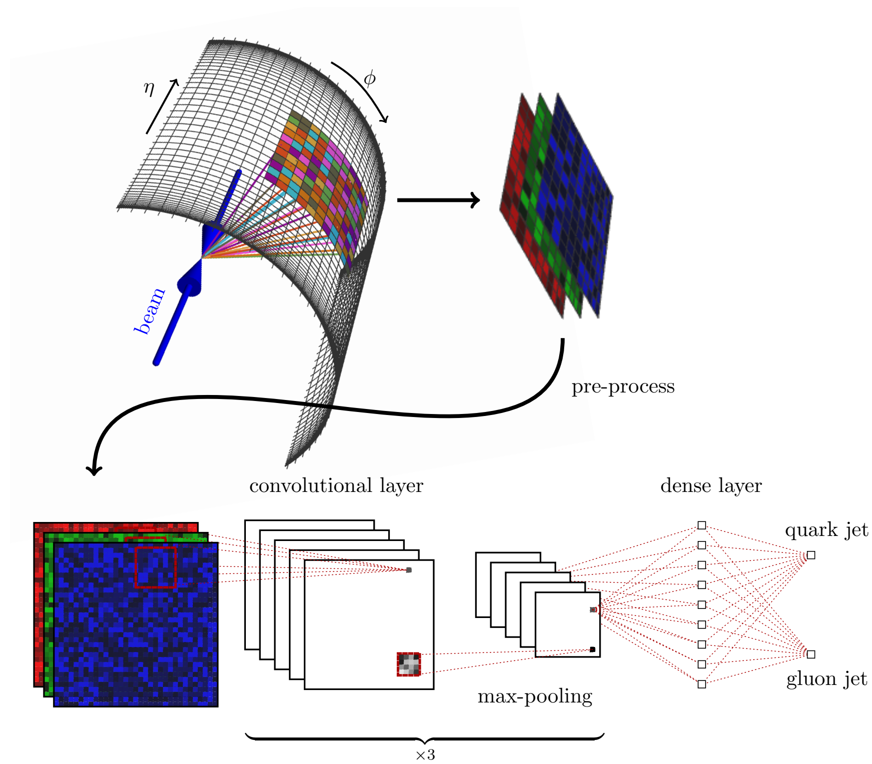
University of California San Diego
Department of Physics
Physics 139/239, Fall 2025
Machine Learning in Physics (4 units)



Instructor: Javier Duarte, jduarte@ucsd.edu, OH by appt. MYR-A 5516
Teaching assistant: Daniel Primosch, dprimosc@ucsd.edu, OH Tu 4–5:00pm MYR-A 5516

Course webpage, Zoom link to lectures:

Canvas: https://canvas.ucsd.edu/courses/56257
Webpage: https://jduarte.physics.ucsd.edu/phys139_239
All assignments will be due through Gradescope (accessed through Canvas).

Course information: This course is an upper-division undergraduate course and introductory graduate course on machine learning in physics. No previous machine learning knowledge is necessary. However, some basic knowledge of calculus, linear algebra, statistics, and Python programming may be expected/useful.

The course structure will consist of weekly lectures on conceptual topics, e.g. statistics, linear algebra, scientific data set exploration, feature engineering, (stochastic) gradient descent, neural networks, and unsupervised learning. Students will learn key concepts in data science and machine learning, including selecting and preprocessing data, designing machine learning models, evaluating model performance, and relating model inputs and outputs to the underlying physics concepts. We will apply these methods to the domains of collider physics, neutrino physics, astronomy, and potentially others. There will be 4 homework assignments. There will also be a final project in which students will work in groups to reproduce the results of an ML in physics research article. A midterm assignment to propose the project will also be required.

Schedule:

Lecture	MWF	12–12:50p	RWAC 0121, Zoom 96709322126
Final exam	Th 12/11/2025	11:30a-2:29p	RWAC 0121, Zoom 96709322126

First lecture: F 9/26/2025
Pre-course survey: Please fill out this pre-course survey and attest that you have done it on Canvas by 9/26/2025.

Textbook: There is no required textbook for this course. At the end of the syllabus, we list a bibliography of (mostly free) textbooks and online resources we will draw from.

Student learning outcomes: Upon successful completion of Physics 139/239, students will be able to:

- Find, explore, select, and preprocess scientific data
- Choose and design machine learning models
- Evaluate model performance and compare to standard benchmarks
- Debug machine learning workflows
- Relate model inputs and outputs to underlying physics concepts
- Collaborate with peers to tackle complex, realistic problems
- Present findings

Grading policy: Your final course grade will be determined according to the following:

- 50% Homework.
- 10% Participation and attendance.
- 20% Midterm: Written proposal for group project.
- 20% Final: Written group project summary, presentation, self/peer evaluations, and code.

Drop policy: The lowest homework score is dropped automatically. This drop policy is designed to account for any illnesses, family, medical, mental, or other emergencies.

If you have an extended emergency (e.g., a long hospital stay) that hinders your ability to turn complete assignments beyond the emergency policy allowance, contact the professor directly as soon as the situation arises.

Discussion board: We will use Slack: ucsdphys139.slack.com

Homework: Each homework will consist of a set of conceptual and programming problems. The assignments will be submitted as Jupyter notebooks or GitHub repositories.

There will be a first deadline (on Fridays at 8:00pm) to submit the homework, which will be graded based on effort and completeness.

There will be a second deadline (on Wednesdays at 8:00pm) to submit corrections for the homework, which will be graded based on effort and correctness. You are required to submit corrections for all assignments even if you believe everything is correct. In that case, for each problem, you should indicate that you've checked the solution and your solution is equivalent.

Midterm and final project: For the final project, students will work in groups of $\sim$4 to reproduce or extend the results of an ML in physics research article. Some candidate articles are listed at the end of the syllabus. The final project deliverables are: (1) a 4-page paper on the project, (2) code provided as a public GitHub repository, (3) a 10-minute presentation by all members of the group during finals week, and (4) self and peer evaluations for group contributions. Students will also be required to submit a 1-page written proposal for the project in Week 7. This is to ensure the project is feasible and to receive feedback from the instructors.

Attendance (lectures): In-person lecture attendance is strongly recommended and worth 10% of your grade. To record your attendance, write your name on the whiteboard or chalkboard at the beginning of lecture. The full 10% will be awarded for attending 80% of the lectures. The lecture hours will be split into conceptual and hands-on portions, with interactive problem-solving and pair programming throughout. Please, bring a laptop that you can program with to lecture. If you do not have one, please contact the instructor, and we will help you. These sessions will be recorded.

Exit tickets: At the end of each class, you will be invited to fill out an exit ticket.

Academic integrity (including AI tools, e.g. ChatGPT): Please read the College Policies section of the UCSD's Policy on Integrity of Scholarship. These rules will be enforced. Cheating includes, but is not limited to: submitting another person's work as your own, copying from any person/source, and using any unauthorized materials or aids during exams.

We understand that many of you may utilize AI tools such as ChatGPT to assist with your coursework. While these tools can be valuable resources for learning and exploring concepts, it is important to use them responsibly. Copying and pasting AI-generated answers without making an effort to understand the material not only undermines your learning but also violates the principles of academic integrity. Although we can often identify the use of such tools, we recognize that some instances may go undetected. Relying solely on these tools without comprehension will hinder your long-term academic and professional development. We encourage you to use AI assistance to enhance your understanding and support your problem-solving processes, but ensure that all submitted work reflects your own effort and comprehension.

Copying from an online solution, a peer's solution, a Chegg solution, or shared work (on Slack, for example) is considered cheating. Collaboration is encouraged, but by the time you start writing your own solution to turn in, you should not be looking at any other source. You should know the rough outline of the solution well enough that you do not need to reference something line-by-line. Plagiarizing a solution but changing variable names is considered cheating. Soliciting help online via Chegg, Quora, etc. is considered cheating. If suspected, you might be asked to rework similar problems in a Zoom one-on-one meeting with the instructor and/or TA.

Any questions on what constitutes an academic integrity violation should be addressed to the instructor; any violation of academic integrity will result in immediate reporting to the UCSD Office of Academic Integrity, and can result in an automatic "F" for the course at the discretion of the instructor.

Counseling and Psychological Services (CAPS): The mission of CAPS is to promote the personal, social, and emotional growth of students. Many services are available to UCSD students including individual, couples, and family counseling, groups, workshops, and forums, consultations and outreach, psychiatry, and peer education. To make an appointment, call (858) 534-755. For more information, visit https://wellness.ucsd.edu/caps/.

Schedule (Subject to change):

Week 0

Friday 9/26: Lecture 01: Course overview, introduction to ML, linear regression; Homework 1 released

Week 1

Monday 9/29: Lecture 02: Introduction to ML, linear regression (cont.)

Wednesday 10/1: Lecture 03: Over/underfitting, bias-variance tradeoff, cross validation

Friday 10/3: Lecture 04: Perceptron learning algorithm, (stochastic) gradient descent; Hands-on: Python/Jupyter, NumPy, Git, debugging

Week 2

Monday 10/6: Lecture 05: Support vector machine

Wednesday 10/8: Lecture 06: Regularization, logistic regression

Friday 10/10: Homework 1 due; Lecture 07: (Boosted) decision trees

Week 3

Monday 10/13: Lecture 08: (Boosted) decision trees (cont.); Hands-on: Scikit-learn, XGBoost, classifying Higgs boson events

Wednesday 10/15: Homework 1 (corrections) due; Homework 2 released; Lecture 09: (Deep) neural networks, backpropagation

Friday 10/17: Lecture 10: Classification metrics, confusion matrix, ROC curve, AUC

Week 4

Monday 10/20: Lecture 11: (Deep) neural networks (cont.), training issue, data standardization; Hands-on: Keras, classifying jets with high-level features

Wednesday 10/22: Lecture 12: Optimizers: (Nesterov) momentum, RMSProp, Adam, skip connections, regularization: dropout, early stopping

Friday 10/24: Homework 2 due; Lecture 13: Types of data, inductive bias, image-like data, convolutional neural networks

Week 5

Monday 10/27: Lecture 14: Convolutional neural networks (cont.)

Wednesday 10/29: Homework 2 (corrections) due; Homework 3 released; Lecture 15: Spherical convolutional neural networks

Friday 10/31: Hands-on: Keras, classifying astronomical data (images)

Week 6

Monday 11/3: Lecture 16: Time-series data, recurrent neural networks

Wednesday 11/5: Lecture 17: Recurrent neural networks (cont.)

Friday 11/7: Homework 3 due; Hands-on: Identifying radio signals (time series)

Week 7

Monday 11/10: Lecture 18: Point cloud and graph-like data, relational inductive bias, permutation invariance/equivariance, graph neural networks and transformers

Wednesday 11/12: Homework 3 (corrections) due; Lecture 19: Graph neural networks and transformers (cont.)

Friday 11/14: Project proposal due; Hands-on: Graph neural networks and transformers, N-body simulations, springs

Week 8

Monday 11/17: Lecture 21: Unsupervised learning, clustering

Wednesday 11/19: Lecture 22: Autoencoders, variational autoencoders

Friday 11/21: Homework 4 due; Lecture 23: Generative modeling; Hands-on: Finding anomalies in LHC/LIGO data

Week 9

Monday 11/24: Lecture 24: Model compression, pruning

Wednesday 11/26: Homework 4 (corrections) due; Lecture 25: Quantization, knowledge distillation; Hands-on: TensorFlow Model Optimization, QKeras

Week 10

Monday 12/1: Lecture 26: Large language models and foundation models

Wednesday 12/3: Lecture 27: Large language models and foundation models (cont.); Hands-on: Fine-tuning LLMs

Friday 12/5: Guest lecture: TBD

Finals Week

Thursday 12/11: Final presentations and projects due

Bibliography:

Textbooks:

[1] Y. S. Abu-Mostafa et al., *Learning from data*, Note: Good general introduction to machine learning (AMLBook, 2012), https://amlbook.com/.

[2] Z. Ivezic et al., *Statistics, data mining, and machine learning in astronomy: a practical python guide for the analysis of survey data*, Note: machine learning applications in astronomy (Princeton University Press, 2014).

[3] I. Goodfellow et al., *Deep learning*, Note: Comprehensive, free textbook on deep learning (MIT Press, 2016), http://www.deeplearningbook.org.

[4] P. Mehta et al., "A high-bias, low-variance introduction to machine learning for physicists", Phys. Rept. **810**, 1 (2019), doi:10.1016/j.physrep.2019.03.001, arXiv:1803.08823, Note: Free on arXiv and oriented at physicists.

[5] M. Erdmann et al., *Deep learning for physics research*, Note: Intermediate deep learning for physics research with Jupyter notebook exercises (World Scientific, 2021), doi:10.1142/12294, http://deeplearningphysics.org/.

[6] F. Chollet, *Deep Learning with Python*, 2nd ed., Note: Good reference by the author of Keras. Free e-book from UCSD Library (Manning, 2021), https://www.manning.com/books/deep-learning-with-python-second-edition.

[7] P. Calafiura et al., *Artificial intelligence for high energy physics*, Note: Mostly reviews of AI applications in high energy physics. Some chapters can be found for free on arXiv. (World Scientific, 2022), doi:10.1142/12200.

Videos:

[8] 3Blue1brown, *But what is a neural network? | Chapter 1, deep learning*, 2017, https://www.youtube.com/watch?v=aircAruvnKk.

[9] 3Blue1Brown, *Gradient descent, how neural networks learn | Chapter 2, deep learning*, 2017, https://www.youtube.com/watch?v=IHZwWFHWa-w.

Reviews:

[10] G. Carleo et al., "Machine learning and the physical sciences", Rev. Mod. Phys. **91**, 045002 (2019), doi:10.1103/RevModPhys.91.045002, arXiv:1903.10563.

[11] HEP ML Community, *A Living Review of Machine Learning for Particle Physics*, 2021, arXiv:2102.02770, https://iml-wg.github.io/HEPML-LivingReview/.

[12] K. Cranmer, U. Seljak, and K. Terao, "Machine Learning" Ch. 41 in Particle Data Group et al., "Review of particle physics", Phys. Rev. D **110**, 030001 (2024), doi:10.1103/PhysRevD.110.030001, https://pdg.lbl.gov/2024/reviews/rpp2024-rev-machine-learning.pdf.

Candidate articles for final project:

[13] A. Aurisano et al., "A Convolutional Neural Network Neutrino Event Classifier", JINST **11**, P09001 (2016), doi:10.1088/1748-0221/11/09/P09001, arXiv:1604.01444.

[14] D. Guest et al., "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks", Phys. Rev. D **94**, 112002 (2016), doi:10.1103/PhysRevD.94.112002, arXiv:1607.08633.

[15] L. de Oliveira et al., "Jet-images — deep learning edition", JHEP **07**, 069 (2016), doi:10.1007/JHEP07(2016)069, arXiv:1511.05190.

[16] P. T. Komiske et al., "Deep learning in color: towards automated quark/gluon jet discrimination", JHEP **01**, 110 (2017), doi:10.1007/JHEP01(2017)110, arXiv:1612.01551.

[17] M. Erdmann et al., "Classification and Recovery of Radio Signals from Cosmic Ray Induced Air Showers with Deep Learning", JINST **14**, P04005 (2019), doi:10.1088/1748-0221/14/04/P04005, arXiv:1901.04079.

[18] A. Khan et al., "Deep learning at scale for the construction of galaxy catalogs in the Dark Energy Survey", Phys. Lett. B **795**, 248 (2019), doi:10.1016/j.physletb.2019.06.009, arXiv:1812.02183.

[19] R. Zhou et al., "Deep ugrizY imaging and DEEP2/3 spectroscopy: a photometric redshift testbed for LSST and public release of data from the DEEP3 galaxy redshift survey", Mon. Notices Royal Astron. Soc. **488**, 4565 (2019), doi:10.1093/mnras/stz1866, arXiv:1903.08174.

[20] LSST Dark Energy Science, Transient, Variable Stars Science Collaboration, "Models and Simulations for the Photometric LSST Astronomical Time Series Classification Challenge (PLAsTiCC)", Publ. Astron. Soc. Pac. **131**, 094501 (2019), doi:10.1088/1538-3873/ab26f1, arXiv:1903.11756.

[21] E. A. Moreno et al., "Interaction networks for the identification of boosted $H \to b\bar{b}$ decays", Phys. Rev. D **102**, 012010 (2020), doi:10.1103/PhysRevD.102.012010, arXiv:1909.12285.

[22] R. Ormiston et al., "Noise Reduction in Gravitational-wave Data via Deep Learning", Phys. Rev. Res. **2**, 033066 (2020), doi:10.1103/PhysRevResearch.2.033066, arXiv:2005.06534.

[23] E. A. Moreno et al., "Source-agnostic gravitational-wave detection with recurrent autoencoders", Mach. Learn.: Sci. Technol. **3**, 025001 (2022), doi:10.1088/2632-2153/ac5435, arXiv:2107.12698.

[24] Majorana Collaboration, "Majorana Demonstrator Data Release for AI/ML Applications", Preprint, 2023, arXiv:2308.10856.

[25] C. Li et al., "Accelerating Resonance Searches via Signature-Oriented Pre-training", Preprint, 2024, arXiv:2405.12972.

[26] S. Miao et al., "Locality-Sensitive Hashing-Based Efficient Point Transformer with Applications in High-Energy Physics", in 41st International Conference on Machine Learning, Vol. 235 (May 2024), p. 35546, arXiv:2402.12535, https://proceedings.mlr.press/v235/miao24b.html.

[27] J. T. Fry et al., "TIDMAD: Time Series Dataset for Discovering Dark Matter with AI Denoising", (2024), arXiv:2406.04378.

Public datasets:

[28] AstroDave et al., *Galaxy Zoo - The Galaxy Challenge*, Kaggle, 2013, https://kaggle.com/competitions/galaxy-zoo-the-galaxy-challenge.

[29] T. Allam et al., *PLAsTiCC Astronomical Classification Challenge*, 2018, https://kaggle.com/competitions/plasticc.

[30] Anaderi et al., *TrackML Particle Tracking Challenge*, 2018, https://kaggle.com/competitions/trackml-particle-identification.

[31] CMS Collaboration et al., *Sample with jet, track and secondary vertex properties for Hbb tagging ML studies (HiggsToBBNTuple_HiggsToBB_QCD_RunII_13TeV_MC)*, CERN Open Data Portal, 2019, doi:10.7483/OPENDATA.CMS.JGJX.MS7Q.

[32] G. Kasieczka et al., *Top quark tagging reference dataset*, version v0 (2018_03_27), 2019, doi:10.5281/zenodo.2603256, https://doi.org/10.5281/zenodo.2603256.

[33] C. Messenger et al., *G2Net Gravitational Wave Detection*, 2021, https://kaggle.com/competitions/g2net-gravitational-wave-detection.

[34] H. Qu et al., *Jetclass: a large-scale dataset for deep learning in jet physics*, version 1.0.0, 2022, doi:10.5281/zenodo.6619768, https://doi.org/10.5281/zenodo.6619768.

[35] *Majorana Demonstrator Data Release for AI/ML Applications*, version 1.0 (Zenodo, 2023), doi:10.5281/zenodo.8257027.

[36] C. Li et al., *JetClass-II*, 2025, doi:10.57967/hf/6336, https://huggingface.co/datasets/jet-universe/jetclass2.