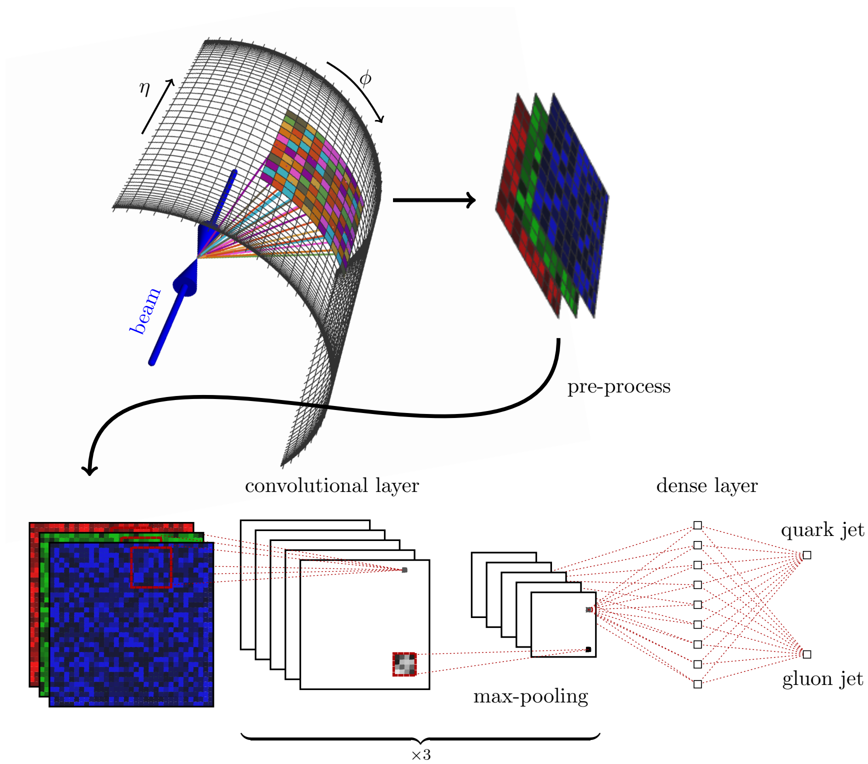
University of California San Diego
Department of Physics
Physics 139/239, Winter 2023
Machine Learning in Physics (4 units)



Instructor: Javier Duarte, jduarte@ucsd.edu, OH TuTh 2:00–3:00pm, Zoom 99550747356
Teaching assistant: Xiaoche Wang, xiw067@ucsd.edu, OH Th 5:30–6:30pm, Zoom 94686360343
Course webpage, Zoom link to lectures:
Canvas: https://canvas.ucsd.edu/courses/41966
Webpage: https://jduarte.physics.ucsd.edu/phys139_239
All assignments will be due through Canvas and 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	TuTh	12:30p–1:50p	SOLIS 109, Zoom 95725188091
Final exam	Tu 3/21/2023	11:30a-2:29p	Location TBD, Zoom TBD

First lecture: Tu 1/10/2023

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 in class, via Slack, and completion of exit tickets.
- 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 and all 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 5:00pm) to submit a "draft" version of the homework, which will be graded based on effort.

There will be a second deadline (on Wednesdays at 5:00pm) to submit a "final" version of the homework, which will be graded based on effort and correctness.

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 20-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 2-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 not required, but strongly recommended. 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: 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.

For homework assignments, copying from an online solution, a peer's solution, a Chegg solution, or shared work (on Discord, 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 1

Tuesday 1/10: Lecture: Course overview, introduction to ML, linear regression, over/underfitting, bias-variance tradeoff, cross validation

Wednesday 1/11: Homework 1 released

Thursday 1/12: Lecture: Perceptron learning algorithm, (stochastic) gradient descent Hands-on: Python/Jupyter, NumPy, Git, debugging

Week 2

Tuesday 1/17: Lecture: Support vector machine, regularization, logistic regression

Thursday 1/19: Lecture: (Boosted) decision trees, Hands-on: Scikit-learn, XGBoost, classifying Higgs boson events

Friday 1/20: Homework 1 (draft) due

Week 3

Tuesday 1/24: Lecture: (Deep) neural networks, backpropagation, training issues

Wednesday 1/25: Homework 1 (final) due; Homework 2 released

Thursday 1/26: Lecture: Data standardization, optimizers: (Nesterov) momentum, RMSProp, Adam, skip connections, regularization: dropout, early stopping, Hands-on: Keras, Classifying jets with high-level features

Week 4

Tuesday 1/31: Lecture: Types of data, inductive bias, image-like data, convolutional neural networks

Thursday 2/1: Lecture: Convolutional neural networks (cont.), Hands-on: Keras, classifying astronomical data (images)

Week 5

Monday 2/6: Homework 2 (draft) due

Tuesday 2/7: Lecture: Time-series data, recurrent neural networks

Thursday 2/9: Lecture: Recurrent neural networks (cont.), Hands-on: Identifying radio signals (time series)

Friday 2/10: Homework 2 (final) due; Homework 3 released

Week 6

Tuesday 2/14: Lecture: Point cloud and graph-like data, relational inductive bias, permutation invariance/equivariance, graph neural networks

Thursday 2/16: Lecture: Graph neural networks (cont.), Hands-on: Spektral, N-body simulations, springs

Friday 2/17: Homework 3 (draft) due

Week 7

Tuesday 2/21: Lecture: Unsupervised learning, clustering, autoencoders

Wednesday 2/22: Homework 3 (final) due; Homework 4 released

Thursday 2/23: Lecture: Variational autoencoders, Hands-on: Finding anomalies in LHC/LIGO data

Friday 2/24: Project proposal due

Week 8

Tuesday 2/28: Lecture: Model compression, pruning, quantization

Thursday 3/2: Lecture: Knowledge distillation, Hands-on: TensorFlow Model Optimization, QKeras

Friday 3/3: Homework 4 (draft) due

Week 9

Tuesday 3/7: Guest lecture: Generative modeling by Dr. Benjamin Nachman (Lawrence Berkeley National Laboratory)

Wednesday 3/8: Homework 4 (final) due

Thursday 3/9: Guest lecture: Equivariant neural networks by Professor Rose Yu (UC San Diego)

Week 10

Tuesday 3/14: Guest lecture: TBD

Thursday 3/16: Guest lecture: Physics-informed neural networks by Dr. Amir Gholami (UC Berkeley)

Finals Week

Tuesday 3/21: Final project 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] 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.

[4] 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/.

[5] 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.

[6] 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:

[7] 3Blue1brown, *But what is a neural network? | Chapter 1, deep learning*, 2017, https://www.youtube.com/watch?v=aircAruvnKk.

[8] 3Blue1Brown, *Gradient descent, how neural networks learn | Chapter 2, deep learning*, 2017, https://www.youtube.com/watch?v=IHZwWFHWa-w.

Reviews:

[9] 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.

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

Candidate articles for final project:

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

Public datasets:

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

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

[23] 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.

[24] 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.

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