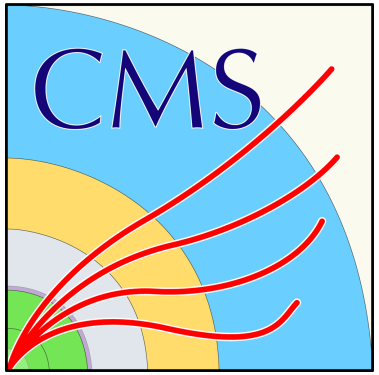
EDUCATION AND ENGAGEMENT OF ML SKILLS

Savannah Thais

Snowmass Computing Frontier Kickoff

08/11/2020



Framing

ML is its own robust and rapidly evolving field. We must engage the community, understand best practices, and actively exchange ideas.

This talk focuses on 4 key areas:

- Knowledge and training
- Outreach and community building
- Open/benchmark datasets
- Ethics and safety of AI

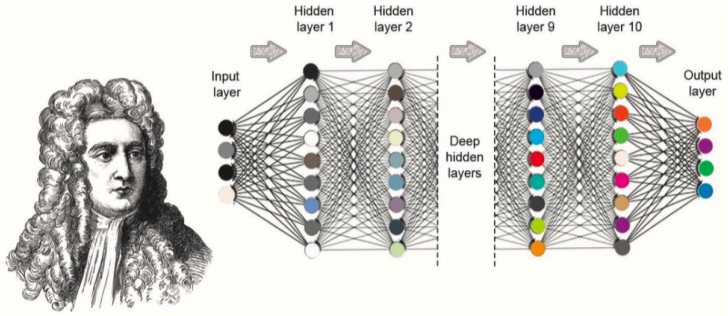


Figure 2. Newton and the machine. Image of sir Isaac Newton alongside a schematic of a 10-layer deep neural network. In each layer (apart from the input layer), a node takes the weighted input from the previous layer's nodes (plus a bias) and then applies an activation function before passing data to the next node. The weights (and bias) are free parameters which are updated during training.

*Caveat: I've always worked in LHC physics, so the existing resources highlighted here are definitely biased towards that. None of this (goals, resources, questions) are meant to be extensive!

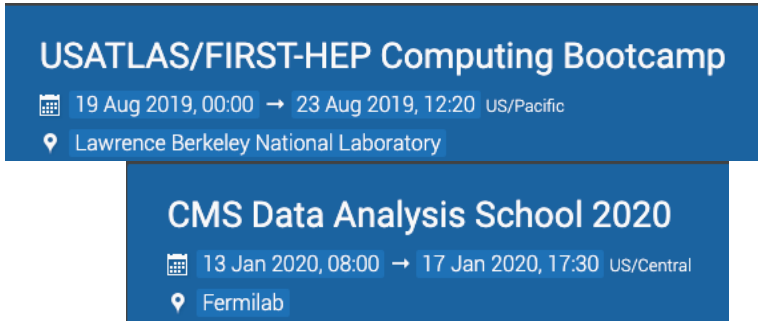
Knowledge and Training

Goals

- Develop a robust understanding of ML within the physics community
- Cultivate robust and transferable technical skills
  - Especially among young physicists
- Provide physics specific ML training materials
  - How techniques are used for our problems
- Create reusable/reproducible education materials

Resources: Summer Schools + Technical Trainings

MLHEP School
Weeklong training eveloped and sponsored by Yandex with collaboration from physicists (HEP focused)

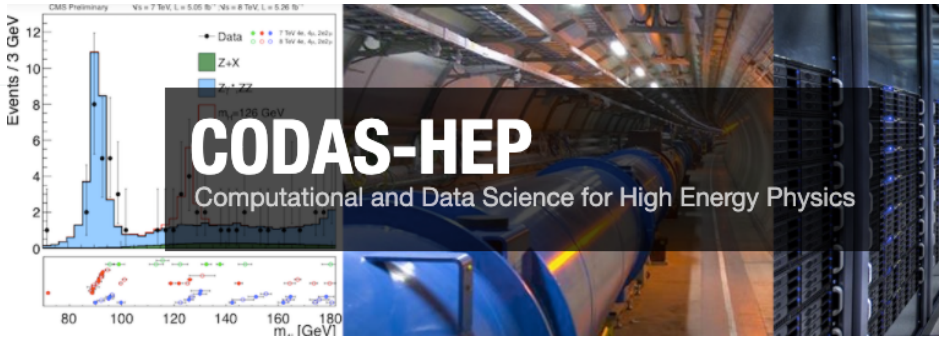


CMS and ATLAS Computing Trainings
Both had short sessions (3-4 hours) on ML, led by experts in the experiments (Experiment focused)

HSF
HEP Software Foundation

The HEP Software Foundation facilitates cooperation and common efforts in High Energy Physics software and computing internationally.

HEP Software Foundation Trainings
Materials for reproducible trainings, ML module in development (LHC focused)



Deep Learning for Science School

Webinar Series: July - September, 2020

Lawrence Berkeley National Laboratory, Berkeley, CA

Deep Learning for Science School
5 day training from ML experts and discussion on use in scientific research (General Science focused)

CoDaS-HEP School
5 day advanced computing training including ~2 days on ML led by domain experts with physics connections (HEP focused)

Resources: Within Experiments and Universities

PY 895 Machine Learning for Physicists. Fall 2019

BU Physics

This is the website for PY 895, Machine Learning for Physicists This website will be updated with HWs and suggested readings. In the fall, the class will be based on our new review A high-bias, low-variance introduction to Machine Learning for physicsits. The review can be downloaded from the arXiv or Physics Reports if for some reason you prefer that formatting. The Jupyter Notebooks can be downloaded from Github.

PHYS 398 MLA - DATA ANALYSIS & MACHINE LEARNING APPLICATIONS

WEB PAGE

https://courses.physics.illinois.edu/phys398MLA

The Grainger College of Engineering Physics

Notes on Contemporary Machine Learning for Physicists

Jared Kaplan

Department of Physics and Astronomy, Johns Hopkins University

DEPARTMENT OF PHYSICS

Short Course on Machine Learning for Physicists

USC Viterbi

AME 599 Machine Learning and Computational Physics
Units: 4
Spring 2020—Monday/Wednesday—11-12:50

Machine Learning in Physics and Astronomy

Kartheik Iyer, John Wu, Raghav Kunnawalkam Elayavalli
Rutgers University
SSPAR Oct 5th 2017

University Courses/Workshops

- Varied in home department and target students
- Generally start with intro to ML then delve into physics applications
- Courses usually require programming homework and extensive pre-reqs

Experiments and Labs

- Provide presentation opportunities for experiment specific uses of ML
- Centralize resources (trainings, papers, etc) for groups
- Support researchers in development of ML methods in their individual experiment

STATUS AND GENERAL NEWS

Fermilab Machine Learning Group

Neutrino Physics and Machine Learning

IML

Inter-Experimental LHC Machine Learning Working Group

About IML | Forum (mailing list) | Meetings

Questions to Consider

- What should be included in graduate curriculums?
- How can experiments/groups fill the gaps?
- What role do summer schools and workshops play?
- How can schools with established programs knowledge share?
- How can we create centralized resources?
- What role should ML experts play in development of the above?

Key Collaborations: CommF2 (Career Pipeline & Development), CommF4 (Physics Education), SEC

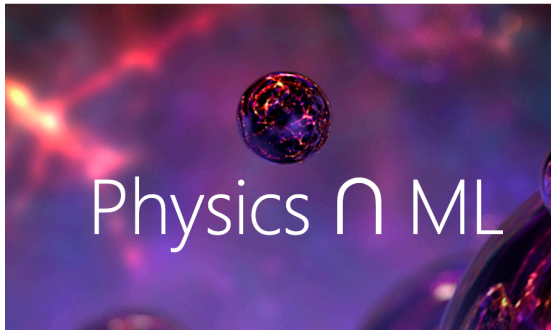
Outreach and Community Building

Goals

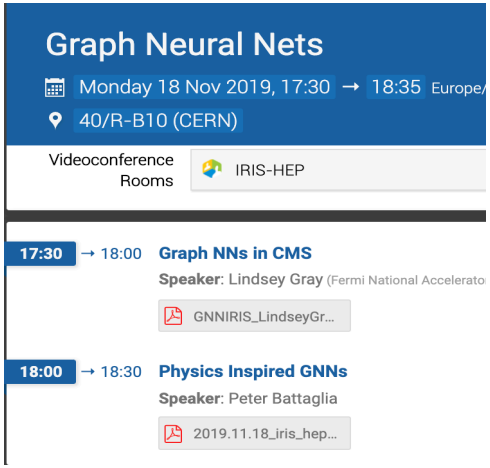
- Develop effective collaborations with industry and Computer Science ML researchers
- Stay up-to-date on new ML methods and results
  - And help develop/inform them
- Share physics-focused results in critical ML venues
- Provide enhanced training and development opportunities
  - Can include internships etc

Current Activities

Microsoft | Research Research areas



Microsoft sponsored



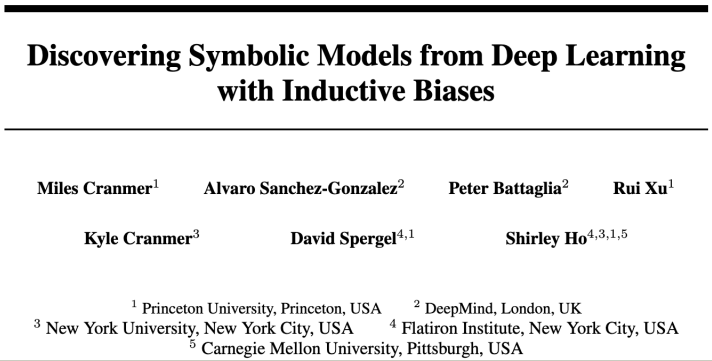
IRIS-HEP sponsored

Deep Learning for Physics

IAS/Princeton sponsored

Lectures Series

Various groups are organizing lectures where ML researchers present to physics groups (and vice versa)



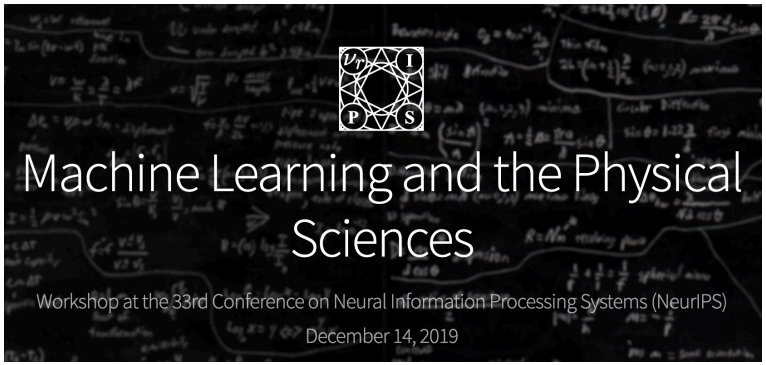
Discovering Symbolic Models from Deep Learning with Inductive Biases

Miles Cranmer[1] Alvaro Sanchez-Gonzalez[2] Peter Battaglia[2] Rui Xu[1]
Kyle Cranmer[3] David Spergel[4,1] Shirley Ho[4,3,1,5]

[1] Princeton University, Princeton, USA [2] DeepMind, London, UK
[3] New York University, New York City, USA [4] Flatiron Institute, New York City, USA
[5] Carnegie Mellon University, Pittsburgh, USA

Individual Collaborations

Many physicists are collaborating directly with industry researchers; some hold dual appointments across physics and CS



Events at ML Conferences

ML and Physical Sciences Workshop at NeurIPS (2017, 2019), resubmitted for 2020. Brings physics applications and physics for ML together.

Questions to Consider

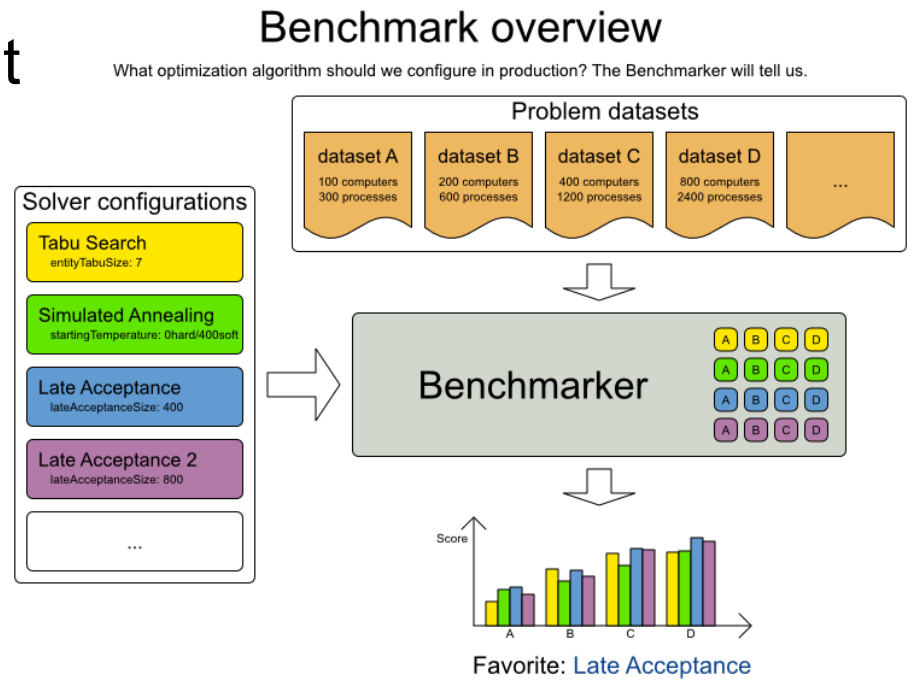
- What are best practices for collaborating across disciplines?
  - Should it originate from the experiments? Universities? Individuals?
- How can we assist physicists in connecting with established ML researchers?
- How can the unique constraints of our problems/experiments inform the development of new ML methods?
- How can we highlight the work physicists are doing to the wider ML community?

Key Collaborations: CommF1 (Applications and Industry), CommF5 (Public Education and Outreach), SEC

Open/Benchmark Datasets

Goals

- Curate datasets to allow easy and accurate comparison of new ML techniques on established physics problems
- Provide open data that is accessible to non-experiment specific researchers
  - Helpful to interested ML researchers and theorists
- Cultivate a culture of open science and reproducible results



Current Efforts

Kaggle Challenges

Allow anyone to submit solutions to simplified physics problems.

IML

Simplified datasets for benchmarking:

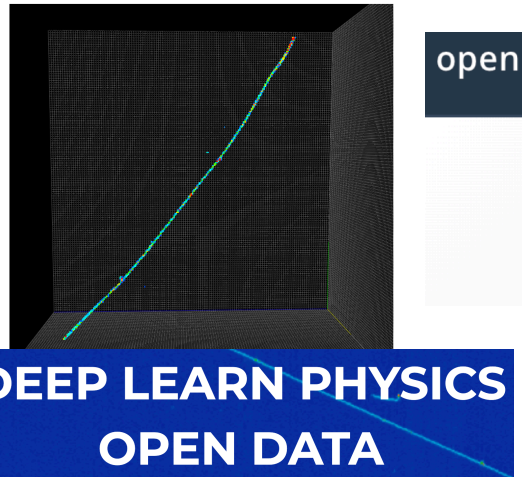
- Top tagging without heavy flavour & pileup: Data and details of arXiv:1707.08966
- Jet substructure: Data from arXiv:16107.08633 at UC Irvines page MLPhysics
- Flavour tagging without pileup: Data from arXiv:1603.09349 at UC Irvines page MLPhysics

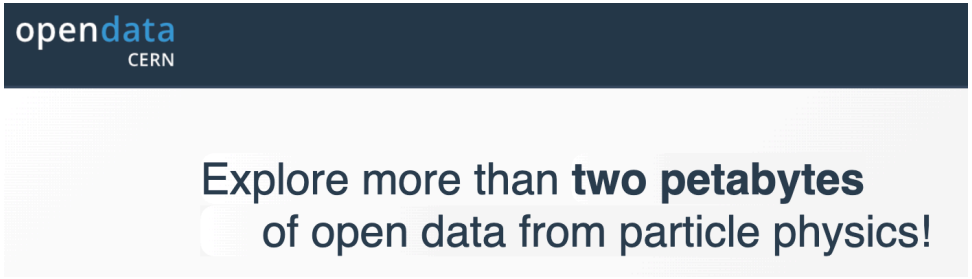
Benchmarking Efforts

IML has started collecting existing benchmark datasets. CMS ML is working on developing more.

Open Datasets

Many experiments are providing open access datasets with necessary scale factors etc





Questions to Consider

- What HEP problems need benchmark datasets?
- Who should create and maintain them?
- How can we ensure open datasets are accessible by non-physicists while still producing useful results?
- How can we engage the broader ML community with these datasets?

Key Collaborations: CommF1 (Applications & Industry), CommF4 (Physics Education), CompF7 (Reinterpretation and long-term preservation of data and code)

Ethics and Safety of AI

Ethics and Safety of AI

Artificial intelligence / Machine learning

Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

Biased policing is made worse by errors in pre-crime algorithms

TECH FIX

Your Doorbell Camera Spied on You. Now What?

curity cameras have gaping security
t yourself.

Machine Bias

s software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

Police mass face recognition in the US will net innocent people

SCIENCE & TECH

epfakes Are Getting Better, Easier to Make, and Cheaper

tHub is becoming a destination site for make-your-own-deepfake software.

COMPUTING

Racial Bias Found in a Major Health Care Risk Algorithm

Black patients lose out on critical care when systems equate health needs with costs

Clearview's Facial Recognition App Has Been Used By The Justice Department, ICE, Macy's, Walmart, And The NBA

Finance

The Apple Card Is Sexist. Blaming the Algorithm Is Proof.

Apple's algorithm seems to be granting women a fractio
borrowing limits. It's a troubling example of machine lea

One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority

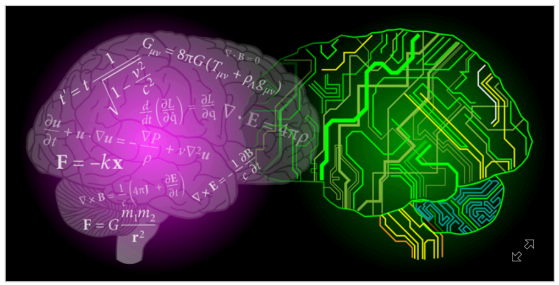
In a major ethical leap for the tech world, Chinese start-ups have built algorithms that the government uses to track members of a largely Muslim minority group.

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process

Why It Matters for Physicists

- **We are community members and developers**
  - Physicists are contributing to and driving ML research
  - Many students who leave the field continue in industry ML
  - We are humans and citizens, these issues affect us and the communities we work and live in
- **We have critical skills and experience**
  - Physics research relies on critical inquiry and robust analysis; we can bring these principles to ML experiments.
  - Our analyses require precise estimations of uncertainties and systematics; these are crucial calculations for publicly deployed ML
  - We seek to understand WHY an algorithm works for a certain problem; these methods can inform algorithmic interpretability
  - Our field has a rich history of public outreach and engagement; we can help develop these programs in ML



Physicists Must Engage with AI Ethics, Now

July 9, 2020 • *Physics* 13, 107

Questions to Consider

- How can we enable inter-disciplinary conversations around these topics?
- How can we ensure the educational resources we develop appropriately address these issues?
- What unique perspectives and skills can physicists contribute to addressing these issues?
- How can we all support our communities and use our technical skills in an equitable and just manner?
- How can we be conscious of our computational power usage and limit it where appropriate?

Key Collaborations: CommF1 (Applications & Industry), CommF3 (Diversity & Inclusion), CommF4 (Physics Education), CommF5(Public Education and Outreach), CommF6 (Public Policy & Government Engagement), SEC

Conclusions

- Machine Learning is becoming inextricably linked to physics research
- We must ensure our community continues to develop these critical technical skills
- We have opportunities to directly contribute to ML research AND to gain from ML researcher involvement
- We must be conscious of the ethical implications of this work