Machine Learning in HEP
Sergei Gleyzer
University of Florida

ACAT 2017
Outline

• Questions
• Challenges
• Inter-Experimental Efforts
ML in HEP Today

Machine learning already at forefront of what we do in HEP

- Physics object identification
- Event type classification
- Object properties regression

Expanding quickly to more areas
Questions
Questions

Can we fully exploit the detectors?

• Raw data, low-level variables

Images: D. Whiteson, K. Cranmer
“End-to-end learning”

- By-passing traditional reconstruction

Talk by M. Andrews
Posing the Problem

How well can we map ML algorithms to HEP problems?

\[ u_1 = f(W_1 x + \theta_1) \quad u_2 = f(W_2 u_1 + \theta_2) \quad u_3 = f(W_3 u_2 + \theta_3) \quad u_4 = f(W_4 u_3 + \theta_4) \]
Defining the Problem

If a problem can be expressed as a known problem

• Apply existing algorithms

If a problem has not been solved

• Push the knowledge boundary forward
If a challenge can be stated as an image or video problem

- Excellent **computer vision** techniques
  - Convolutional Neural Networks
  - Examples:
    - Jet and calorimeter images, neutrino images
- High degree of **success** in HEP
Examples

Jet images with convolutional nets

L. de Oliveira et al., 2015
Examples

Neutrinos with convolutional nets

Aurisiano et al. 2016

NOvA Preliminary

76% Purity
73% Efficiency

An equivalent increased exposure of 30%
Examples

Tracking with recurrent nets (LSTM)

Time dimension (state memory)

HEPTrkX Project, talk by A. Tsaris
Meaningful Physics

Can we extract features with meaningful physics?

• from low-level variables

Are we able to understand ML models

• physics interpretations
Pile-up with CNN

What is learned?

- Train a single $4 \times 4$ filter and inspect it.
- Pixel-wise: $p_T^{N, LV} \approx p_T^{N, tot} - \frac{1}{2} p_T^{C, PU}$
- This is linear cleansing with $\gamma_0 = 2/3!$

$$p_T^{N, LV} = p_T^{N, tot} + \left(1 - \frac{1}{\gamma_0}\right) p_T^{C, PU}$$

PUMML Talk by E. Metodiev
Deep learning improvements apply to regression as well.
Defining the Problem

How to best use domain knowledge we have accumulated?

• in designing the algorithms

Strings  
Jet Images  
Jet Clustering

Krefl, 2017  
de Oliveira et al., 2015  
Louppe et al., 2017
Uncertainties

- Decision making
Uncertainties matter

Multi-Task Model

See Kyle’s talk
Multi-Task Model

Example: flavor tagging (talk by M. Stoye)

Regression examples (talks by L. de Oliveira and Lu Wang)
Can we do ML in real-time?

- ML: live video analysis, medical, self-driving cars
- HEP Trigger Systems (software and hardware)
  - See talks by A. Carnes and S. Skambraks
Other Applications

Unsupervised learning (no labels)
  • Anomaly detection, unexpected physics

Generative models
  • Simulation and better training

Optimization and tuning
  • Bayesian optimization etc.
All very exciting directions
• with many challenges to overcome
Opportunity to re-examine how we have done things until now
• try to do more with less
A bit of history

There was a point in time where people went away from neural networks

- **Learning problems** with back-propagation and over-training
- **Overcome** in 2006 by Hinton et al.
Computation

After that current success relies on significant brute-force computation

• GPUs, no magic sauce

Sustaining and achieving greater progress requires computational resources

• Well-justified by physics returns
HEP-ML Community
Community

Past:
• Pockets of great ideas and applications

Present:
• Organized, open discussion platforms, cross-fertilization of ideas, more collaborative, inter-experimental

Recent Results: speak for themselves
Today’s Community

- Open source
- Reproducible science
- Collaborative vs. secretive
HSF effort on S&C on HL-LHC timescale

- **Examine** where we are today and where we are headed (more in *Liz’s talk today*)
- **Machine learning** is one of the areas
  - Many contributors over the past 6 months
Community White Paper

Discussion of many of the topics in this talk

• Applications and R&D
• Available resources and needs
• Collaboration with ML community
  – Collaborative datasets
Summary

Machine learning already front and center

• **Significant progress** over the past two years, efforts going from R&D to physics results

• Significant challenges remain in interpretability, scalability and real-time inference
  – Exciting areas in data science
  – Growing community – become part of it.
Thank You