Machine Learning to Statistical Criteria

Charanjit Kaur Khosa

University of Manchester

HEP Seminar, Oklahoma State University
4th August, 2022
Outline

• Why to use Machine Learning in Particle Physics?

• How to implement these methods with statistical approaches?

• What are the current developments in this direction?

• What is a way forward?
Why Physics Beyond the Standard Model?

**Dark Matter**

[CERN courier](https://www.cern.ch)

**Neutrino Masses**

[Symmetry Magazine](https://symmetrymagazine.org)

**Flavor Puzzle**

[CERN Courier](https://www.cern.ch)

**Baryon Asymmetry**

[Symmetry Magazine](https://symmetrymagazine.org)

**Hierarchy Problem?**

**Gravity? …**
Current Status of BSM

Absence of direct signal for New Physics might be due to:

Nature of BSM

- New Physics signature is beyond the reach of the current experiments
- Known BSM models do not include the “correct” model

How do we search for it?

- Traditional analysis strategies are not suitable
- Model dependent searches (pre-bias)

Need to go beyond these limitations
A New Perspective

Large Hadron Collider (LHC)
Big data brings challenges
1. Large volumes of data
2. High dimensionality of the data sets
3. Large number of model parameters

Neutrino Experiments
Challenges that require new techniques
1. Data reconstruction
2. Physics inference
3. Physics modelling
4. Uncertainty quantification
What Is Machine Learning?

A. Machine learning is a subdomain of artificial intelligence.

B. It uses statistical models and algorithms to identify patterns in and/or fit data without using explicit instructions.
ARTIFICIAL INTELLIGENCE

Emergence
Deep learning in the trigger
The most stable tetraquark yet

Welcome to the digital edition of the September/October 2021 issue of *CERN Courier*.

As data volumes surge, deep learning is becoming increasingly important in particle physics. This special edition on artificial intelligence (AI) captures two new trends: using “unsupervised” deep learning to spot anomalous events, and designing AI that can “think not link”. Community-organised data challenges are leading the way (p27) and deep learning could even be used in the level-one triggers of LHC experiments (p31). To keep up with the cutting edge of AI research, physicists are reaching out to computer science and industry (p36): the latest developments could help explore theory space (p51) and build trust in AI to do more of the heavy lifting throughout the analysis chain (p49). We also explore recent thinking that an ordered simplicity may emerge from the complexity of deep learning in a similar way to statistical mechanics and quantum field theory (p39).

Elsewhere in the issue: a tribute to Steven Weinberg (p65); a SciFi upgrade for LHCb (p43); reports from the summer conferences (p19); the most stable tetraquark yet (p7); quantum gravity in the Vatican (p59); anisotropies point to cosmic-ray origins (p11); and much more.

To sign up to the new-issue alert, please visit: [http://comms.iop.org/k/iop/cerncourier](http://comms.iop.org/k/iop/cerncourier)

To subscribe to the magazine, please visit: [https://cerncourier.com/p/about-cern-courier](https://cerncourier.com/p/about-cern-courier)
# Machine Learning at the Energy and Intensity Frontiers

A. Radovic et al., Nature 560 (2018) no. 7716,41

## Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Years of data collection</th>
<th>Sensitivity without machine learning</th>
<th>Sensitivity with machine learning</th>
<th>Ratio of P values</th>
<th>Additional data required</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS$^{24}$</td>
<td>2011-2012</td>
<td>$2.2\sigma$, $P = 0.014$</td>
<td>$2.7\sigma$, $P = 0.0035$</td>
<td>4.0</td>
<td>51%</td>
</tr>
<tr>
<td>$H \rightarrow \gamma\gamma$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATLAS$^{43}$</td>
<td>2011-2012</td>
<td>$2.5\sigma$, $P = 0.0062$</td>
<td>$3.4\sigma$, $P = 0.00034$</td>
<td>18</td>
<td>85%</td>
</tr>
<tr>
<td>$H \rightarrow \tau^+\tau^-$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATLAS$^{99}$</td>
<td>2011-2012</td>
<td>$1.9\sigma$, $P = 0.029$</td>
<td>$2.5\sigma$, $P = 0.0062$</td>
<td>4.7</td>
<td>73%</td>
</tr>
<tr>
<td>$VH \rightarrow b\bar{b}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATLAS$^{41}$</td>
<td>2015-2016</td>
<td>$2.8\sigma$, $P = 0.0026$</td>
<td>$3.0\sigma$, $P = 0.00135$</td>
<td>1.9</td>
<td>15%</td>
</tr>
<tr>
<td>$VH \rightarrow b\bar{b}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS$^{100}$</td>
<td>2011-2012</td>
<td>$1.4\sigma$, $P = 0.081$</td>
<td>$2.1\sigma$, $P = 0.018$</td>
<td>4.5</td>
<td>125%</td>
</tr>
<tr>
<td>$VH \rightarrow b\bar{b}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Early Successes

Complete list of references: HEPML-LivingReview

**H → b¯b decay**

- Boosted H → b¯b Significance
  - Machine Learning
  - Standard Search

**Top Tagging**

Machine Learning techniques increase the discovery potential of the experiments
Commonly Used ML Methods
Neural Network (NN)

Basic Structure

A. Input layer nodes: set of observables/kinematical features/images

B. Number of hidden layers (shallow or deep NN)

C. Output layer: predictions

Schematic of a Neural Network

Train the network using training sample and make predictions for the test (real) dataset
Convolutional Neural Networks (CNN)

Input image

Convolutions

Pooling

Fully connected

Convolution layer

Pooling

Fully connected

Source: stanford.edu
Jet Images as an Input to CNN

Energy of charged particles
Energy of neutral particles
Number of charged particles

- Center
- Crop
- Normalize

P. T. Komiske, E. M. Metodiev and M. D. Schwartz, JHEP 01(2017), 110
Benchmark 1: Top and QCD Jets

the input dataset

SM $t\bar{t}$ and QCD diJet production, $\sqrt{s} = 13$ TeV

Madgraph + Pythia

Leading jet with $p_t > 750$ GeV, $R=1$ Anti-kt jet

$\Delta\eta = 0.087, \Delta\phi = 0.087$

Averaged over 50K events
CNN for Top Vs QCD Classification (2C Baseline Classification)

100K Events (balanced data)
Training: Test data = 70:30%

Cross-Entropy Loss function

Batch Size = 100
Epochs = 100

Input 25x25

Conv2D 30, 3X3, Stride=1

Conv2D 30, 3X3, Stride=1, padding=1

Conv2D 40, 3X3, Stride=1, padding=1
Maxpooling 2

Conv2D 40, 3X3, Stride=1, padding=1
Maxpooling 2
Dropouts = 0.3

Linear Layer 300

Predictions (Nc)

Related work
G. Kasieczka, T. Plehn, M. Russell and T. Schell, JHEP 05 (2017), 006
S. Macaluso and D. Shih, JHEP 10(2018), 121
Auto-Encoders

Can be used as anomaly detector:
1. Train with the background sample.
2. Compare how the reconstructed output is different from the input (reconstructed error).

Reconstructed error will be more for the anomalous event

M. Farina, Y. Nakai and D. Shih, arXiv: 1808.08992, any others (see HEPML-LivingReview)
Statistics: Key Terms

• Probability Model ($p(x, y | \mu, \theta)$): Term used for the statistical model.

• Likelihood $L(\mu, \theta)$: Value of the probability for a fixed data set as a function of the parameters.

• Hypotheses (null ($H_0$) or alternative ($H_1$)): The null hypothesis is what one tries to test and the alternative hypothesis is what data is giving us.

• Test Statistics ($t_\mu$): measures how well data favours either hypothesis.

• Significance: Quantify the discrepancy of observed data with the null hypothesis.
Standard Practice in Particle Physics

- **Discovery Significance**: discrepancy of the observed data with the null hypothesis. An actual experiment is conducted for which the null hypothesis predicts the probability of the observed data.

- **Expected Significance**: One considered signal hypothesis instead of actual data which predicts the probability of the observed outcome.

- **Limit Setting**: Role of null hypothesis and signal hypothesis is reversed. One rejects a parameter value if the discrepancy of the observed data with the signal hypothesis is larger than a threshold value.
Statistical Tests

Test statistic, profile likelihood and significance

\[ t_\mu = -2 \ln \lambda(\mu) \]
\[ \lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})} \]
\[ Z = \sqrt{t_\mu} \]

\[ L(\mu) = \prod_{j=1}^{N} \frac{(\mu_j s_j + b_j)^{n_j}}{n_j!} e^{-(\mu_j s_j + b_j)} \]

\( s_j \) and \( b_j \) are the expected number of signal and background events in the \( j^{th} \) bin.

\( n_j \) is the total number of events in the \( j^{th} \) bin according to the alternate hypothesis \( H_1 \).

Expected discovery significance

\[ E[Z_0] = \sqrt{-2 \left[ \sum_{j=1}^{N} \left( s_j + (b_j + s_j) \ln \left( \frac{b_j}{b_j + s_j} \right) \right) \right]} \]

Expected exclusion significance

\[ E[Z_e] = \sqrt{2 \left[ \sum_{j=1}^{N} \left( s_j + b_j \ln \left( \frac{b_j}{b_j + s_j} \right) \right) \right]} \]

Our Approach

• A simple dictionary connecting the dots of this known knowledge about machine learning methods and statistical approaches.

• More like thinking about possible ways to connect the dots.

• An attempt to build a recipe for any type of output to statistical significance.

Directly use ROC?

Toy experiments and log-likelihood?

Should we use full distribution?

What are the equivalent options for unsupervised learning?

How to implement uncertainties?

\[
\frac{S}{\sqrt{B}}
\]

Asimov Significance

Good First Approximation

- Choose a point on the ROC curve
- Calculate $S/\sqrt{B}$ for that point considering the cross-section
Log-likelihood Ratio (LLR) Test Statistics

Likelihood for hypothesis H in a single pseudo-experiment

\[ L_H = \prod_{i=1}^{\text{events}} pdf_H(x) \]

QCD vs top+QCD

\[ \Lambda_{QCD} = -2 \ln \frac{L(QCD, QCD)}{L(top + QCD, QCD)} \]

\[ \Lambda_{top+QCD} = -2 \ln \frac{L(QCD, top + QCD)}{L(top + QCD, top + QCD)} \]

PDF and LLR

QCD vs QCD+Top

PDF

P(Top)

10^1
10^0
10^{-1}

0.0 0.2 0.4 0.6 0.8 1.0

QCD vs QCD + Top, P_{cut(top)} = 0

N_{qcd} = 97658
N_{top} = 106

Frequency

LLR

-15 -10 -5 0 5 10 15

QCD + TOP
QCD
QCD
Translating to Significance

\[ H_0 \text{ (Null Hypothesis)} \quad H_1 \text{ (Alternative Hypothesis)} \]

Type II Error  \( \beta \)  Type I Error  \( \alpha \)

\[
\begin{align*}
\alpha &= \frac{\int_{-\infty}^{\Lambda_{cut}} f_0(\Lambda) d\Lambda}{\int_{-\infty}^{\infty} f_0(\Lambda) d\Lambda} \\
\beta &= \frac{\int_{-\infty}^{\Lambda_{cut}} f_1(\Lambda) d\Lambda}{\int_{-\infty}^{\infty} f_1(\Lambda) d\Lambda}
\end{align*}
\]
Significance Comparison

QCD vs QCD + Top, $P_{\text{cut(top)}} = 0$

$N_{\text{qcd}} = 97658$
$N_{\text{top}} = 106$

Frequency

LLR

$P_{\text{cut(top)}} = 0.0$

Significance

$N_{\text{top}}$

Legend:
- $a = 1.35 \times 10^{-3}$
- $a = 2.87 \times 10^{-7}$
- $a = \text{Symmetric}$
- $z = \frac{S}{\sqrt{B}}$
- $z = \frac{S}{\sqrt{B} + \bar{B}}$
- $z = \text{As mov}$
A cut on Classifier Output

QCD vs QCD+Top

PDF

$P_{\text{cut(top)}} = 0.5$

$N_{\text{Top}}$

Significance

$\alpha = 1.35 \times 10^{-3}$

$\alpha = 2.87 \times 10^{-7}$

$\alpha = \text{Symmetric}$

$\alpha = \text{Asymmetric}$

$Z = S/\sqrt{B}$

$Z = S/\sqrt{S + B}$

$Z = \text{Asimov}$
For $P_{\text{cut(top)}} = 0.0$ and $P_{\text{cut(top)}} = 0.5$, the significance curves are plotted against $N_{\text{Top}}$.

- Black dotted line: $\alpha = 1.35 \times 10^{-3}$
- Black solid line: $\alpha = 2.87 \times 10^{-7}$
- Red dotted line: Symmetric $\alpha$
- Red solid line: Asymmetric $\alpha$
- Green dotted line: $Z = S/\sqrt{B}$
- Blue dotted line: $Z = S/\sqrt{S+B}$
- Yellow dotted line: $Z = \text{Asimov}$
Effect of Uncertainties

$P_{\text{cut}}(\text{top}) = 0.0$

![Graph showing the effect of uncertainties on significance vs. $N_{\text{Top}}$.]
Effect of Uncertainties

QCD jet image before and after blurring.
Higgs EFT Benchmark: Supervised Approach

\[ \mathcal{L}_{EFT} = \mathcal{L}_{SM} + \mathcal{L}_{BSM} \quad \text{where} \quad \mathcal{L}_{BSM} = \frac{1}{\Lambda^{2n}} \sum_{i} c_i \mathcal{O}_i \]

\[ \mathcal{L}_{BSM} \supset i g \frac{c_{HW}}{m_W^2} (D^\mu H) \dagger \sigma_a (D^\nu H) W_\mu^a \]

\[ p_T^{b_1}, p_T^{b_2}, p_T^{l_1}, p_T^{l_2}, p_T^H, \eta_H, \phi_H, \delta R_{l_1 l_2}, \delta R_{b_1 l_1}, M_T^Z, p_T^Z, \delta \phi_{l_1 b_1}, d \phi_{l_1 l_2} \]
Higgs EFT Benchmark

DNN output

SM vs EFT

PDF

P(EFT)

Significance

N_{EFT}

\( a = 1.35 \times 10^{-3} \)

\( a = 2.87 \times 10^{-3} \)

\( a = \) Symmetric

\( a = \) Asymmetric

\( Z = \) Asimov
Unsupervised Machine Learning

Reconstruction error using VAE
Other Interesting Proposals

• Madminer and DNNlikelihood (inference from likelihood ratios)


• Anomaly Detection


• And many more (see HEPML-LivingReview for a complete list of references.)
Summary and Outlook

- ML techniques are emerging as a competitive tool to look for new phenomena in complex data and more efficiently identify known objects. It seems like an appropriate tool for particle physics.

- Our community is adapting these techniques for various tasks. I talked about the ways to bridge the gap between the ML outputs and statistical criteria.

- In particular, I talked about how from the ML output one could perform a simple hypothesis testing.

- As a next step, it would be interesting to consider more complicated situations to assess the performance of this set-up.

- Other ways of using a neural network as a test statistic and further uncertainty quantification.
Future public releases of likelihoods from experiments could incorporate information on ML training outputs following the lines of this paper.

Publishing statistical models: Getting the most out of particle physics experiments

Abstract

The statistical models used to derive the results of experimental analyses are of incredible scientific value and are essential information for analysis preservation and reuse. In this paper, we make the scientific case for systematically publishing the full statistical models and discuss the technical developments that make this practical. By means of a variety of physics cases — including parton distribution functions, Higgs boson measurements, effective field theory interpretations, direct searches for new physics, heavy flavor physics, direct dark matter detection, world averages, and beyond the Standard Model global fits — we illustrate how detailed information on the statistical modelling can enhance the short- and long-term impact of experimental results.