Deep Learning and the Future of Scientific Discovery

BSPS 2024, York

July 18, 2024

Martin King

Munich Center for Mathematical Philosophy, LMU Munich

Introduction

Two major difficulties for understanding

1. Opacity of the network
2. Uninterpretability of the output

Introduction

Two major difficulties for understanding

1. Opacity of the network
2. Uninterpretability of the output

- However, DNNs are a powerful tool to aid in scientific discovery
  - as such, they will help us understand the world

Overview

Machine Learning in HEP

- Around for decades (BDTs, multivariate analysis)[1]
- DNNs are changing the game

[1] (Albertsson et al., 2019; Bourilkov, 2020)

"In the relatively few years that modern machine learning [deep learning] has existed, it has already made traditional collider physics obsolete. In the past, physicists, including me, would devote their efforts to understanding signatures of particular particles or processes from first-principles: why should a stream of pions coming from a W boson decay look different than a stream coming from an energetic gluon? Now we simply simulate the events, and let neural network learn to tell the two samples apart."[2]

[2](Schwartz, 2021, p. 10)

"Our analysis shows that recent advances in deep learning techniques may lift these limitations by automatically discovering powerful non-linear feature combinations and providing better discrimination power than current classifiers—even when aided by manually-constructed features,"[3].

[3](Baldi, 2014, p. 9)

- Reduced role for physics knowledge (leading-order processes)
- High-performance decisions may not based on physicist identified features
  - Do we understand these decisions?

Many roles:

- distinguishing dark matter signatures in LHC physics [4], in searching for exotic Higgs decays [5], and in jet flavour tagging [6], top tagging [7], optimizing the reduction of a nuisance parameter [8], to improve the triggers at the LHC [9], anomaly detection [10], and many more examples are emerging every week.

[4](Khosa et al., 2021)
[5](Jung et al., 2022)
[6](Munoz et al., 2022)
[7](Kasieczka et al., 2019)
[8](D'Agnolo and Wulzer, 2019)
[9](Pol et al., 2020)
[10](Collins et al., 2018; Pol et al., 2020; Chekanov and Hopkins, 2022)

Anomaly Hunting: AEN

Autoencoder Networks

- Consist of an encoder and a decoder
  - Transform inputs into low-dimensional latent representations (e.g. abstract vector space of feature values)
  - Then elaborate the latent representations back to high-dimensional representations
  - Trained to minimize the error, calculated as the difference between the output and the input
    - unsupervised (or better, self-supervised)

Anomaly Hunting: AEN

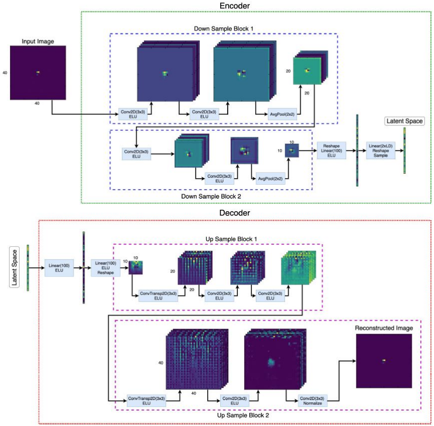

Figure: (Fraser et al., 2022)

Basic Programme

1. Preprocess data to be suitable for the network
2. Train and optimize network
   - Have it learn the SM background with small reconstruction error
3. Perform BSM benchmarking
   - Ensure that the model gives large reconstruction errors for a variety of BSM scenarios, for additional W', Z', leptoquarks, charged Higgs, etc.
4. Test on real CERN data
5. Study the flagged regions with various other resources

Signal over Background

- Learn background and subtract it from the real data, leaving a cleaner signal (if there is one)

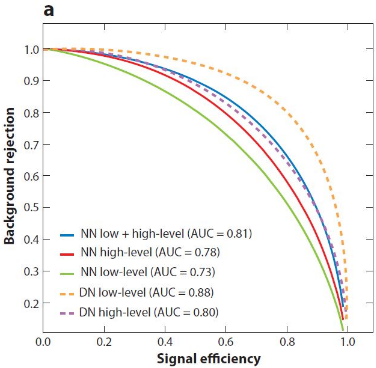

Figure: (Guest et al., 2018)

Reducing Dimensionality

Steps

1. reconstruction algorithms are used to process raw data into objects (clusters and tracks)
2. use this to estimate the energy and momentum of particles
3. identify particles
4. build *event-level* summaries
5. perform event selection for further analysis

The reconstruction and selection are traditionally based on *physicist-identified* features of the data

- DNNs outperform at every step

Classification

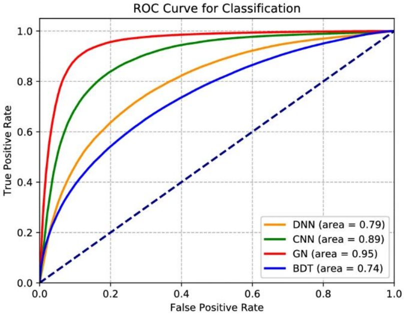

Figure: γ vs. π^0 (Belayneh et al., 2020)

End-to-End

- Perform high-level event classification on low-level data
- Use 4-momenta directly, or images of angular distributions, without explicitly resolving particles in intermediate steps [11]

[11] (Andrews et al., 2020; Baldi et al., 2022; Farina et al., 2020)

End-to-End

"The shallow neural networks and BDTs trained with the high-level features perform significantly better than those trained on only the low-level features, demonstrating the importance of feature engineering in shallow machine learning models. . . only the deep learning approach shows nearly equal performance using the low-level features and the complete features. This suggests that it is automatically discovering high-level abstractions similar to those captured by the hand-engineered features, obviating the need for laborious feature engineering,"[12]

[12](Baldi et al., 2022, p. 6–7)

Opacity Worries

- We can't see what the network has learned
- We don't understand why the network works so well
- We don't understand how each decision is made

Uninterpretability Worries

- Outputs can be non-linear correlations between huge number of variables
- Variables do not necessarily correspond to measurable quantities and are not always easily represented or described

Not Understanding Worries

These are real concerns, but they shouldn't make us worry that they block or don't contribute to our understanding

Not Understanding Worries

These are real concerns, but they shouldn't make us worry that they block or don't contribute to our understanding

- Science uses extremely complex non-DNN models

Not Understanding Worries

These are real concerns, but they shouldn't make us worry that they block or don't contribute to our understanding

- Science uses extremely complex non-DNN models
- DNNs are only a part of a multi-pronged approach of discovery

Not Understanding Worries

These are real concerns, but they shouldn't make us worry that they block or don't contribute to our understanding

- Science uses extremely complex non-DNN models
- DNNs are only a part of a multi-pronged approach of discovery
- Not like areas where there are ethical issues, one-off decisions

Not Understanding Worries

These are real concerns, but they shouldn't make us worry that they block or don't contribute to our understanding

- Science uses extremely complex non-DNN models
- DNNs are only a part of a multi-pronged approach of discovery
- Not like areas where there are ethical issues, one-off decisions
- We have increasingly better xAI methods

Not Dissimilar

DNNs are not relevantly dissimilar from

- extremely complex models and simulations
  - lots of models and theories are successful, provide explanations, but are not transparently understood

Not Dissimilar

DNNs are not relevantly dissimilar from

- other MI approaches (precision measurements, e.g.)
  - in flagging anomalous data, they indicate where traditional model building and testing can focus

Is Dissimilar

DNNs in HEP are relevantly dissimilar from

- those DNNs making ethical or political decisions (self-driving cars, city planning, legal decisions, etc.)
    - many groups try different approaches, low-cost of failure

Linear Interpretable Model-agnostic Explanation method

- "If the LIME model is influenced by a particular feature, we then infer that the original black box model was too."

[13] (Fleisher, 2022)

xAI LIME

Linear Interpretable Model-agnostic Explanation method

- "If the LIME model is influenced by a particular feature, we then infer that the original black box model was too."
- perturbs inputs and passes them through the DNN

[13] (Fleisher, 2022)

xAI LIME

Linear Interpretable Model-agnostic Explanation method

- "If the LIME model is influenced by a particular feature, we then infer that the original black box model was too."
- perturbs inputs and passes them through the DNN
- it learns a simple mapping, focusing on closeness to original input

[13] (Fleisher, 2022)

xAI LIME

Linear Interpretable Model-agnostic Explanation method

- "If the LIME model is influenced by a particular feature, we then infer that the original black box model was too."
- perturbs inputs and passes them through the DNN
- it learns a simple mapping, focusing on closeness to original input
- generate reports of important features for given decisions

[13] (Fleisher, 2022)

xAI LIME

Linear Interpretable Model-agnostic Explanation method

- "If the LIME model is influenced by a particular feature, we then infer that the original black box model was too."
- perturbs inputs and passes them through the DNN
- it learns a simple mapping, focusing on closeness to original input
- generate reports of important features for given decisions
  - Been argued that we can explain DNNs just like we explain the world[13]

[13] (Fleisher, 2022)

Conclusion

- We have added a powerful tool to our scientific discovery toolbox
- They will aid in discovery and therefore help us better explain and understand the world

Thank you

Martin King
m.king@lmu.de
www.philphys.com
Munich Center for Mathematical Philosophy, LMU Munich

This research was supported by the DFG.

References:

Albertsson, K. et al. (2019). Machine learning in high energy physics community white paper.

Andrews, M., Paulini, M., Gleyzer, S., and Poczos, B. (2020). End-to-end physics event classification with CMS open data: Applying image-based deep learning to detector data for the direct classification of collision events at the LHC. *Computing and Software for Big Science*, 4(1).

Baldi, P., Sadowski, P., and Whiteson, D. (2022). Deep Learning From Four Vectors.

Baldi, P., S. P. . W. D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Commun*, 5:4308.

Belayneh, D., Carminati, F., Farbin, A., Hooberman, B., Khattak, G., Liu, M., Liu, J., Olivito, D., Pacela, V. B., Pierini, M., Schwing, A., Spiropulu, M., Vallecorsa, S., Vlimant, J.-R., Wei, W., and Zhang, M. (2020). Calorimetry with deep learning: particle simulation and reconstruction for collider physics. *The European Physical Journal C*, 80(7).

Bourilkov, D. (2020). Machine and Deep Learning Applications in Particle Physics. *Int. J. Mod. Phys. A*, 34(35):1930019.

Chekanov, S. and Hopkins, W. (2022). Event-Based Anomaly Detection for Searches for New Physics. *Universe*, 8(10):494.

Collins, J., Howe, K., and Nachman, B. (2018). Anomaly detection for resonant new physics with machine learning. *Phys. Rev. Lett.*, 121:241803.

D'Agnolo, R. T. and Wulzer, A. (2019). Learning new physics from a machine. *Phys. Rev. D*, 99:015014.

Farina, M., Nakai, Y., and Shih, D. (2020). Searching for new physics with deep autoencoders. *Phys. Rev. D*, 101:075021.

Fleisher, W. (2022). Understanding, idealization, and explainable ai. *Episteme*, 19:534–560.

Fraser, K., Homiller, S., Mishra, R. K., Ostdiek, B., and Schwartz, M. D. (2022). Challenges for unsupervised anomaly detection in particle physics. *Journal of High Energy Physics*, 2022(3).

Guest, D., Cranmer, K., and Whiteson, D. (2018). Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, 68(1):161–181.

Jung, S., Liu, Z., Wang, L.-T., and Xie, K.-P. (2022). Probing Higgs boson exotic decays at the LHC with machine learning. *Phys. Rev. D*, 105(3):035008.

Kasieczka, G., Plehn, T., Butter, A., Cranmer, K., Debnath, D., Dillon, B. M., Fairbairn, M., Faroughy, D. A., Fedorko, W., Gay, C., Gouskos, L., Kamenik, J. F., Komiske, P. T., Leiss, S., Lister, A., Macaluso, S., Metodiev, E. M., Moore, L., Nachman, B., Nordström, K., Pearkes, J., Qu, H., Rath, Y., Rieger, M., Shih, D., Thompson, J. M., and Varma, S. (2019). The Machine Learning landscape of top taggers. *SciPost Phys.*, 7:014.

Khosa, C. K., Sanz, V., and Soughton, M. (2021). Using machine learning to disentangle LHC signatures of Dark Matter candidates. *SciPost Phys.*, 10(6):151.

Munoz, J. M., Batatia, I., and Ortner, C. (2022). Boost invariant polynomials for efficient jet tagging. *Mach. Learn. Sci. Tech.*, 3(4):04LT05.

Pol, A. A., Berger, V., Cerminara, G., Germain, C., and Pierini, M. (2020). Anomaly Detection With Conditional Variational Autoencoders. In *Eighteenth International Conference on Machine Learning and Applications*.

Schwartz, M. D. (2021). Modern machine learning and particle physics. *Harvard Data Science Review*, 3(2):1–14.