



# Editorial: Mathematical Fundamentals of Machine Learning

David Glickenstein<sup>1</sup>, Keaton Hamm<sup>2\*</sup>, Xiaoming Huo<sup>3</sup>, Yajun Mei<sup>3</sup> and Martin Stoll<sup>4</sup>

<sup>1</sup> Department of Mathematics, University of Arizona, Tucson, AZ, United States, <sup>2</sup> Department of Mathematics, University of Texas at Arlington, Arlington, TX, United States, <sup>3</sup> H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, United States, <sup>4</sup> Chair of Scientific Computing, Department of Mathematics, Technische Universität Chemnitz, Chemnitz, Germany

**Keywords:** machine learning, deep learning, computer vision, interpretability, neural network

## Editorial on the Research Topic

### Mathematical Fundamentals of Machine Learning

With an abundance of data originating from all aspects of life, machine learning, and in particular deep learning, has powered new successes in artificial intelligence. These advances originate from research efforts both in industry and academia, with research spanning fields such as statistics, computer science, optimization, and numerical analysis, but also borrowing from neuroscience and physics among others. While the results are astounding, a deeper understanding of the fundamental principles of machine learning is needed in order to better understand the success and limitations of machine learning techniques.

Among the machine learning techniques that could benefit from more theoretical understanding are deep neural networks, principle component analysis, analysis via coding or thresholding, and various types of network analysis. Application of such techniques to complicated data such as time series pose especially useful and difficult problems. In each case, one might ask when methods translate or generalize to other situations, how much the structure of the data (for instance restriction to a subspace or submanifold) affects the method, and how to choose appropriate parameters to ensure good fit while still avoiding overfitting.

This special issue brings together researchers across disciplinary boundaries that focus on building theoretical foundations and bridging the theoretical gaps in various learning methods. We hope this issue serves as an invitation for researchers to think about fundamental questions from a wide variety of fields, including pure and applied mathematics, statistics and operations research, and computer science and engineering.

The first paper, *Deductron—A Recurrent Neural Network* by Rychlik, studies Recurrent Neural Networks (RNN) by constructing an example of structured data motivated by problems from image-to-text conversion (OCR) of complex scripts. The problem of decoding the script can be considered as a time series which requires long-term memory to decode. An RNN capable of decoding the aforementioned sequences is constructed. The RNN is constructed by inspection, i.e., its weights are guessed by calling a sequence of carefully designed steps and can be compared to how someone would try to learn by hand. The possibility of using simulated annealing and variants of stochastic gradient descent (SGD) methods to train this RNN is also explored.

The second paper, *Statistical Analysis of Multi-Relational Network Recovery* by Wang et al., deals with the problem of recovering a multi-relational network from a small subset of observed edges when the network possesses a low-dimensional latent structure. The authors propose (penalized) maximum likelihood estimators and establish their nearly optimal properties in terms of minimax risk. They also validate their theoretical results via simulation and a real data example in knowledge base completion.

## OPEN ACCESS

### Edited and reviewed by:

Stefan Kunis,  
University of Osnabrück, Germany

### \*Correspondence:

Keaton Hamm  
keaton.hamm@uta.edu

### Specialty section:

This article was submitted to  
Mathematics of Computation and  
Data Science,  
a section of the journal  
Frontiers in Applied Mathematics and  
Statistics

**Received:** 02 March 2021

**Accepted:** 11 March 2021

**Published:** 07 April 2021

### Citation:

Glickenstein D, Hamm K, Huo X,  
Mei Y and Stoll M (2021) Editorial:  
Mathematical Fundamentals of  
Machine Learning.  
Front. Appl. Math. Stat. 7:674785.  
doi: 10.3389/fams.2021.674785

The third paper, *From Learning Gait Signature of Many Individuals to Reconstructing Gait Dynamics of One Single Individual* by Hsieh and Wang, emphasizes data-driven computing paradigms when investigating two interesting problems arising from analysis of wearable sensor data. The first is to differentiate gait signatures of different individuals where named Principle System-State Analysis (PSSA), a data-driven algorithm generalizing principle component analysis, is proposed to represent the gaits. The second is to reconstruct individual gait dynamics in full where two algorithms are developed: an algorithm based on cluster trees called Local-1st-Global-2nd (L1G2) and a landmark computing algorithm. In the literature, the classical ad hoc approaches are to choose a fixed state and then estimate the state's locations throughout the entire time series. The proposed algorithms provide a new approach via data driven algorithms for feature extraction, landmark computing, and fine-scale gait recognition.

The fourth paper, *Understanding Deep Learning: Expected Spanning Dimension and Controlling the Flexibility of Neural Networks* by Berthiaume et al., concerns the power of generalization of neural networks. The ability for neural networks to generalize is fundamental to their success across a wide variety of applications, but is quite challenging to quantify. This article introduces a measure, called the expected spanning dimension (ESD), which quantifies the flexibility that a given neural network has independent of its training data. This measure is then shown to be correlated with testing accuracy across a variety of network architectures and data sets. In particular, ESD can be used to explain the success of certain architectures such as ResNets.

Finally, *Rethinking Breiman's Dilemma in Neural Networks: Phase Transitions of Margin Dynamics* by Zhu et al. considers

margin enlargement of training data, which has been an important strategy for boosting the confidence of training of perceptrons toward good generalizability. Breiman [1] shows a dilemma: a uniform improvement on margin distribution does not necessarily reduce generalization errors. In this paper, the authors revisit Breiman's dilemma in deep neural networks from a novel perspective based on phase transitions of normalized margin distributions in training dynamics. The authors make precise the observation that the expressive power of a neural network model relative to the complexity of the dataset can be observed better by the dynamics of normalized margins in training rather than by counting the number of parameters in the network.

## AUTHOR CONTRIBUTIONS

All authors contributed to the formation, invitation, and editorial work of this Research Topic and editorial article.

## FUNDING

DG and KH were partially supported by NSF CCF-1740858 (University of Arizona Transdisciplinary Research In Principles of Data Science TRIPODS). DG was partially supported by NSF DMS-1760538. XH was partially supported by the Transdisciplinary Research Institute for Advancing Data Science (TRIAD) at Georgia Tech enabled by NSF CCF-1740776 (TRIPODS). XH was partially supported by NSF DMS-2015363. YM was partially supported by NSF DMS-1830344 and DMS-2015405. MS was partially supported by SAB 100378180.

## REFERENCES

1. Breiman L. Prediction games and arcing algorithms. *Neural Comput.* (1999) 11:1493–517.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Glickenstein, Hamm, Huo, Mei and Stoll. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.