2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)

Timothy Doster
Pacific Northwest National Laboratory

Tegan Emerson
Pacific Northwest National Laboratory
Colorado State University
University of Texas, El Paso

Henry Kvinge
Pacific Northwest National Laboratory
University of Washington

Nina Miolane
University of California, Santa Barbara

Mathilde Papillon
University of California, Santa Barbara

Bastian Rieck
Institute of AI for Health, Helmholtz Munich
Technical University of Munich

Sophia Sanborn
University of California, Santa Barbara

Preface

In the last 8 years, deep learning has been applied to an increasing range of tasks, domains, and data types. Each of these problem settings brings its own structure, priors, and symmetries that should be incorporated into the learning pipeline to achieve generalizable and efficient performance. While one may be able to do this in an ad hoc manner in simpler settings, history has shown that in more complex cases, leveraging existing mathematical machinery, which captures the essence of the problem can offer distinct advantages. The exciting puzzle for the ML researcher is then to understand how to transform theory into a suitable learning framework in which we can actually compute.

At the same time, it is well known that the empirical performance of deep learning has far outstripped the theory explaining it, leading to many situations where we know that a specific method works, but we don't know why. The difficulties are compounded by the fact that modern deep learning generally operates on: (i) huge datasets, (ii) that are embedded in a high-dimensional ambient space, and (iii) learned by models that are wildly overparametrized. This leads to situations where no component of the learning system can be comprehensively visualized or naively understood. Mathematics again offers valuable tools here since it has been studying the irreducibly complex and high-dimensional for at least the last 200 years.

The 2nd ICML Workshop on Topology, Algebra, and Geometry in Machine Learning is an exercise in bringing together researchers working in both of the above threads to exchange ideas, present recent work, and form new collaborations. This proceedings collection captures some of the rich flow of ideas that happened in the workshop. It is also a testament to the breadth of ways that mathematics is currently being applied to modern ML, from the topology of models and datasets to equivariance of models to group actions to applications of hyperbolic geometry to learning problems. Worth highlighting is the Topological Deep Learning Challenge that was held during the workshop; this challenge was to foster reproducible research in Topological Deep Learning by crowd-sourcing the open-source implementation of neural networks on topological domains.

The reader will note that we have chosen to focus on a subset of mathematical disciplines: topology, algebra, and geometry. The reason for this is that these areas are both exceptionally deep and well-developed and in our estimation so far under-utilized in deep learning. Indeed, we would point to the explosive growth of the field of equivariant architectures as an example of the potential that ideas from these disciplines have to advance machine learning. Most exciting of all, existing work is only just beginning to scratch the surface of what these fields offer. In this collection we see hints of where the future may take us. Together, these collected works point to an exciting future for applications of pure math to machine learning.

As a final note, the editors of this collection would like to thank the many reviewers who, with their vast collective knowledge, were able to suggest feedback that not only improved the papers themselves, but sometimes even suggested promising new directions of research. Without their help this collection would not have been possible.