

Rethinking the Geometry of Machine Learning: “Physics of Data” Five Years Later

Wednesday, 6 April 2022 17:15 (45 minutes)

In physics the data that is acquired in experiments are highly-controlled and often taken with specific goals in mind. However, the notion of a “Physics of Data” is about using mathematical tools developed in physics to understand data acquired in more open-ended and uncontrolled environments. As the data acquisition process becomes more opaque and distant from any particular purpose, we have to be more careful about our assumptions. We are no longer protected by the remarkable intuition of experimental physicists to find the right thing to measure. Instead, the data is often whatever can be measured about some process in the world and our task is to sift through the resulting mess. The key concept we need is how data and relevance to some specific task interact to create a predictive model.

Note, the data may or may not support the chosen task and different tasks lead to different models even with the same data. This talk will outline how one of the crucial concepts in theoretical physics, fiber bundles or gauge theories, provides a framework for understanding how data, relevance and models are related. For the particular example of computer vision, principal fiber bundles can play a key role since the fibers are Lie groups much like in particle physics. These connections demonstrate the importance of understanding the geometric structure of the data using symmetry transformations well-known to fundamental physics. This may offer a more sensible way forward than the endless tinkering with neural network models in Deep Learning that more often stumble into success and failure without much insight into how the data is actually organized in high-dimensional spaces. Building new tools adapted from physics for this pursuit is an opportunity to fundamentally change how data is analyzed and well-suited to those in the “Physics of Data” Masters program.

Presenter: BYERS, Jeff (Naval Research Laboratory)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications