



Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Some thoughts on Machine Learning

Gabriel Perdue

S2I2 Machine Learning Parallel Session

2017 / May / 2

Machine (deep) learning and HEP - different worries?

- In discovery science we care about some things that are different from what (at least some) in industry care about:
 - We worry a lot about *uncertainty* in algorithm performance.
 - Very often physicists would prefer algorithms that are strictly inferior to others, but where the uncertainty can be better specified (limits to this, of course).
 - Sometime we have models that we know are wrong, and can guess by how much / parameterize our ignorance to some level.
 - Sometimes our models are not even wrong, and parameterizing our ignorance is very difficult. We may be able to put *bounds* on our errors, but estimating the contours may impossible today.
 - The bar for claiming discovery is very high. This is not so unique to us, and is not a new observation, but being able to carefully audit an algorithm and explain its decision-making process is very important.

Simulation

- One of the features HEP may be able to offer is we have very large, very high quality simulation sets.
 - Untold hours of effort have been devoted to making our simulations both very realistic and very detailed.
 - This is a rich playground not only for physics, but for algorithm development - it is possible to take slices of simulation that are very complex and hide and show relationships at a wide variety of levels.
 - And we have a lot of simulation! Plus huge simulated sets from different versions of our physics models...
 - We are very worried about domain differences between our data and our simulation. How do we manage this?
 - There are, of course, a lot of tricks for managing bias in training, but quantifying it is crucial for us.

Discovery science

- What are the sort of analyses and/or questions does AI/ML/DL make possible that can't exist without it?
- What sort of questions are predicated on the existence of ubiquitous, easy to use deep learning?
- Applications might look the way we (HEP folks) “expect” - which is to say, advanced reconstruction and signal discriminants. But this is addressing old questions with new tools at some level. How can we use AI/ML/DL to provide entirely new views on our data?
 - This is especially tricky for HEP because we operate in a frame where we both (a) think we know how to analyze our data well, and (b) have spent many years building a mode and frame for organizing our thoughts.
 - Applications may both be very “mundane” (automated visualization “pretty-ifiers”, management software, etc.) that frees up time to focus on science or very far fetched (AI that analyzes data for us).