DIANA Fellowship Proposal: Instrumenting and Studying Adam and Other Optimization Algorithms in Pytorch

Adam, A Method for Stochastic Optimization, is an algorithm for first-order gradient-based optimization of stochastic objective functions that is widely used in machine learning (ML). In training models for pattern recognition in experimental high energy physics that are qualitatively different from the image processing problems for which Adam was originally designed, it is unclear how its parameters should be optimized for the best learning over thousands of epochs requiring weeks of GPU time. To study this question, we plan to instrument the Adam source code in Pytorch so that it can report what it is doing epoch by epoch. Once the class is instrumented, we will use it to study correlations between the evolution of internal parameters and changes in cost functions for models whose long-term training behavior is already documented.

The incumbent will work under the supervision of a physicist and a computer scientist from the University of Cincinnati who have been developing the targeted machine learning algorithms for several years. Candidates for this position should have a good working knowledge ofi C++, Python, and use of `git`. Experience with Jupyter notebooks and Pytorch is preferred. Experience with other machine learning frameworks, such as TensorFlow or Keras or MatLab, will also be considered favorably. The anticipated duration of the project is the three month period May - July, 2021, although there is some flexibility related to the exact start and finish dates.

Simon Akar (physics) and Gowtham Atluri (computer science) will supervise the student. A timeline with deliverables is provided on the next page.

Timeline

weeks 1-2 learn the existing code for training our models and become familiar with the ADAM source code at the Python and C++ levels.

weeks 3-4 modify the ADAM Python code to allow it to return the learning rate and weight decay parameters at each epoch along with various internal state variables.

weeks 5-6 Using already trained models, re-train using the same intermediate sets of weights as starting points to observe the correlations of the internal ADAM parameters to the evolution of the training and validation set cost-functions.

weeks 7-8 Using the same trained models as in the previous two weeks, study the effects of using alternative input parameters for ADAM.

weeks 9-11 Carefully document the work done earlier in the project for internal use and instrument other optimatizon algorithms (such as ADAMAX) to allow them to report their parameters and internal states similarly. Continue studies from previous two weeks.

weeks 12-13 Document work done for public consumption and present results publicly.

At the end of the project, the student will present his/her work at an IRIS-HEP topical meeting.