Scalable cyberinfrastructure applications

Team: J. Brehmer\textsuperscript{1 2}, K. Cranmer\textsuperscript{1 2}, Irina Espejo\textsuperscript{1}, S. Macaluso\textsuperscript{1 2} and H. Müller\textsuperscript{1}

Institutions: \textsuperscript{1}Center for Data Science, New York University \textsuperscript{2}Department of Physics, New York University

The SCAILFIN Project

---

**MadMiner on REANA**

Simulators provide a high-fidelity description of phenomena. If they are too complex the likelihood is intractable leading to poor statistical analysis.

MadMiner (Machine learning-based inference for particle physics \cite{1}, \cite{2}). Its goal is to estimate parameters using advances in Neural Networks.

To deploy MadMiner in a scalable way we need:

- containerization
- parallelization
- reusability

---

**Excursion**

Goal is to find level sets of black-box functions that are expensive to evaluate. Examples: test statistics from complex simulations.

Evaluate the black box function at interesting points only instead of evaluating at whole regular grid. We use a Gaussian process to: interpolate between samples and model uncertainty in the knowledge of the black box function.

The acquisition function regulates the exploration vs exploitation tradeoff. Select one that minimizes global uncertainty of the location of the excursion set.

Future: efforts will focus on scalability wrt the dimensionality of the function domain space. Example, likelihood ratio as function of mass, charge, spin, ...

---

**ROB Reproducible Open Benchmark Platform**

ROB is an experimental prototype for enabling community benchmarks of data analysis algorithms. The goal of ROB is to allow user communities to evaluate the performance of their different data analysis algorithms in a reproducible competition-style format.

The workflow template and input data are defined by a coordinator. The template contains placeholders for workflow steps that are implemented by the participants (e.g., with Docker containers). The backend processes the submission workflows. The user interface allows participants to submit new runs and to view the results.

---

\cite{1} J. Brehmer, G. Louppe, J. Pavez and K. Cranmer. Constraining Effective Field Theories with Machine Learning, Phys. Rev. Lett. 2018 121, 111801