Integrating MLFlow into AGC Workflow

AGC Demo Day #2 - 24 February, 2023
Elliott Kauffmann
Integrating Machine Learning into AGC

- Provide options for physicists who want to incorporate ML into their analyses on analysis facilities
- Need to address both training and inference
- Tools to utilize: MLFlow, NVIDIA Triton
- AGC Demo Day #1: Triton Inference Server (https://indico.cern.ch/event/1218004/)
- AGC Demo Day #2: MLFlow
Previous AGC Pipeline
AGC Pipeline with ML

Data access → Selection → Normalize and separate events into subsamples → Model fitting and hyperparameter optimization → Select best models for inference → Model storage

Modularity & interfaces → Declarative configuration

Columnar data → Event selection, systematic uncertainties → Machine Learning Inference → Histograms → Statistical model building → Workspace → Statistical analysis → Result, diagnostics
MLFlow

- Track metrics, parameters, and artifacts while running machine learning code
- Organize models/model versions in central repository
- Why would you want to use MLFlow?
  - Monitoring training of complex models which may train over the course of many hours or even days
  - Keeping track of model versions during hyperparameter optimization
- MLFlow + Triton plugin
  - Can move stored models directly from MLFlow model repository to Triton inference server
  - → nicely integrate training and inference pipelines
MLFlow + Model Optimization

- Initialize model with different hyperparameters → need to discover optimal parameters to maximize performance
- Want to parallelize hyperparameter optimization while still keeping track of performance

**MLFLOW EXPERIMENT**

- TRIAL1
- TRIAL2
- TRIAL3
- TRIAL4
- TRIAL5
- TRIAL6
- TRIAL7
- TRIAL8
- TRIAL9
- TRIAL10

**MLFLOW MODEL REGISTRY**

- REGISTERED_MODEL
  - Version1
  - TRIAL8

**TRITON SERVER**

- MODEL_NAME
  - config.pbtxt
  - TRIAL8
MLFlow set-up in analysis facility

User machine learning code + MlflowClient API

KUBERNETES CLUSTER

Metadata Store

Artifact Store
MLFlow + AGC Demonstration

- AGC ML task: assigning jets to parent partons
- Simplified task for the purposes of this demo: signal/background classification using BDT
- [https://github.com/ekauffma/agc-mlflow-demo](https://github.com/ekauffma/agc-mlflow-demo)
- MLFlow link: [https://mlflow.software-dev.ncsa.cloud/#/experiments/10](https://mlflow.software-dev.ncsa.cloud/#/experiments/10)