Challenge: Training the best NNFF model for Aspirin trajectory dataset

Optimize feature setups (either with Behler-Parrinello or GMP, preferably GMP), model structure and training process to get the best NNFF model for the aspirin trajectory dataset in terms of test set MAE.

Dataset: <http://quantum-machine.org/gdml/>

download the aspirin trajectory (scroll down and it’s the fourth one in the yellow box)

Hyper-parameters to optimize

1. Feature setup
   1. Gaussian

G2: list of etas

G2: list of rs\_s

G4: list of etas

G4: list of zetas

G4: list of gammas

Cutoff (don’t go larger than 10, as the molecule is small)

* 1. GMP

List of sigmas (correspond to radial probe, don’t go larger than 5.0 for the largest one)

List of MCSH orders

List of group for each order

1. Model

num\_layers (number of NN layers)

num\_nodes (number of nodes per layer)

1. Optim (training process)

lr (learning rate)

batch\_size

epochs

Additional challenge: try to test your models in parallel on PACE

\*\*Please keep save\_fps to False

PS:

You can use the scripts here as resources: <https://github.com/ray38/GMP_AmpTorch_Tests/tree/main/MD17_examples/BP_comparison>

prepare\_data.py has codes to prepare the training data and

BP\_HDNN\_test.py and GMP\_SNN\_test.py contain feature/model setups you can use to start with