Updates
ConVAE Jets
(20/04/2022)

Breno Orzari, Cauê Evangelista, Jefferson Fialho

SPRACE-ML (track-ml)

Agenda:

- Partial optimization results (JEFF|BRENO)
- VAE with EMD as loss function and partial results (CAUE)
- Partial results testing coupling flows (BRENO)
- Next steps (JEFF|CAUE|BRENO)

Partial optimization results

- Standard study: static hyperparameters set (with no optimization stack)
- Optimization objective function: Minimize the sum(EMD) value
- Standard study hyperparameters set:

```
alpha:                1.0
batch_size:           100
beta:                 0.998
gamma:                1.0
gamma_1:              1.0
gamma_2:              10.0
learning_rate: 0.0001
```

- EMD value from standard study with 300 epochs: **0.0085**
- The best EMD value from optimization step with 300 epochs: **0.0042**

Partial optimization results

- the standard study reached the best optimization value only at the **1500th epoch**
- The hyperparameter set obtained by the optimization converges **5x faster** than the standard study.
- hyperparameter set obtained by the optimization stack:

```
alpha:                0.214241
beta:                 0.996329
gamma:                0.100417
gamma_1:              0.928241
gamma_2:               9.54339
learning_rate: 0.000471193
```

VAE with EMD as loss function and partial results

- Previously:
  - Device error: *solved cloning x_pos and x_decoded_pos*
  - Investigate differentiability: Torch.autograd -> It's differentiable
  - Implementation in the reconstruction loss:
    - CVXPY lib not found, running with QPTH
    - Loss term: average over the batch of jets
    - Multiplicative term added for later optimization: beta_emd
- Partial results:
  - Train loss: 42.24 - epoch 55
  - Validation loss: 42.57 - epoch 55
- Issues:
  - *1 week -> ≈ 50 epochs*

Testing coupling flows

- Using Optuna the best model have EMD = 0.0038
  - Training the model for more epochs to see its behavior
- Need to test using other jets dataset:
  - Gluon jets with more particles (including zero-padded ones)
  - Light quark jets, top jets and W jets with 50 (or 100) particles
- Use other metrics that Raghav also used to evaluate the GAN
  - FPND
  - W1EFP (and the individual EFP plots)
  - W1M
  - W1P

Next steps:

- ~~Run optimization stack with flows~~ -> Still running
- Evaluate ideas:
- log(x) -> histogram(x) -> EMD(x) -> logEMD?
  - Increase the importance of small values
- Using separate EMD values to minimize instead sum(EMD)
  - AVOID the disproportionality from different EMD ranges.
  - 1st experiment:
    - EMD_mass
    - sum(all_other_EMDs)
  - 2nd experiment:
    - EMD_mass
    - sum(EMD_eta, EMD_phi)
    - sum(EMD_pt_EMD_energy)

Next steps:

- Investigate why the training step it's taking so long with EMD.
- Start to implement WaveletEMD
- Evaluate the usage of Jensen-shannon Divergence (JSD)
  - A **symmetrized** and **smoothed** version of the Kullback–Leibler divergence (KLD)
  - handle non-overlapped distributions
  - JSD handles zeros and tail values
- Create a study to evaluate the impact of creating/reconstruction jets using different metrics (EMD, W-EMD, JSD, KLD, GHD)
  - Paper?