Our Approach

Deep neural network with two kinds of predictions:

- Binary classification: muon vs. everything else
- Multi-class classification: muon, pion, kaon, proton, unmatched

Data Preparation

- Taking data from ROOT trees and converted it to pandas DataFrames
- Evenly sampled among classes from different MC generated processes
  - Each process produces kinematically different muons
  - Muons also have different origins
- One-hot encoded the classes for categorical output
- Normalized data

Network Input

- Minimal model: good muon variables (cut-based benchmark)
- Complex model: energy, position, track information (MVA benchmark)
- Custom: combination of the two

Network Architecture

- 4 hidden layers
- 128 neurons/layer
- ReLU activation
- Softmax activation on last layer
- Adam optimizer
- Categorical cross entropy loss

Training Stats

- 35/65 test/training split
- 10/90 validation/training split
- 100 epochs
- 256 batch size
- Accuracy, loss, efficiency, and ROC curve

Results: Binary Classifier

Network statistics

Training

- Accuracy: 0.9916
- Loss: 0.0249

Validation

- Accuracy: 0.9247
- Loss: 0.95241

Test

- Accuracy: 0.9235
- Loss: 0.4939

Results: Binary Classifier

Muon Classification

- Recall (Correct ID rate): 99.9869%
- Purity (MisID rate): 9.2677%

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Purity} = \frac{FP}{TP+FP}$$

Results: Multiclass Classifier

Network statistics

Training

- Accuracy: 0.9303
- Loss: 0.1925

Validation

- Accuracy: 0.6804
- Loss: 1.7296

Test

- Accuracy: 0.6797
- Loss: 1.6903

Results: Multiclass Classifier

Muon Classification

- Recall (Correct ID rate):
- Purity (MisID rate):

Pion Classification

- Recall:
- Purity:

Kaon Classification

- Recall:
- Purity:

Proton Classification

- Recall:
- Purity:

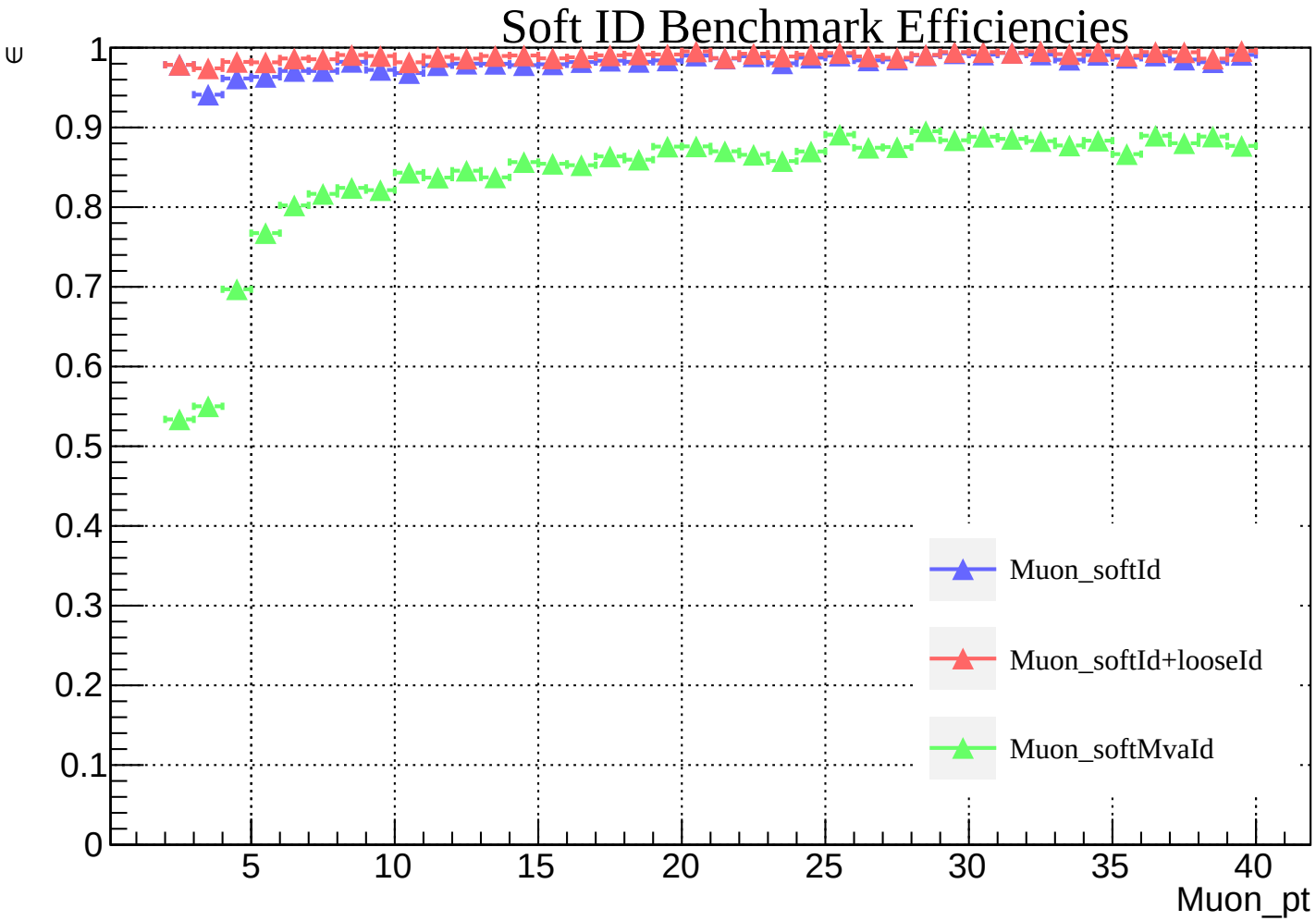
$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Purity} = \frac{FP}{TP+FP}$$

Baseline Model

- Cut-based ID - uses cuts on a few key variables
- Soft MVA - gradient boosted regression forest
- Looked at 3 different samples: DY+Jets, TT+Jets, and QCD
- Efficiency $\epsilon = \#$ true muons that pass ID/$\#$ true muons

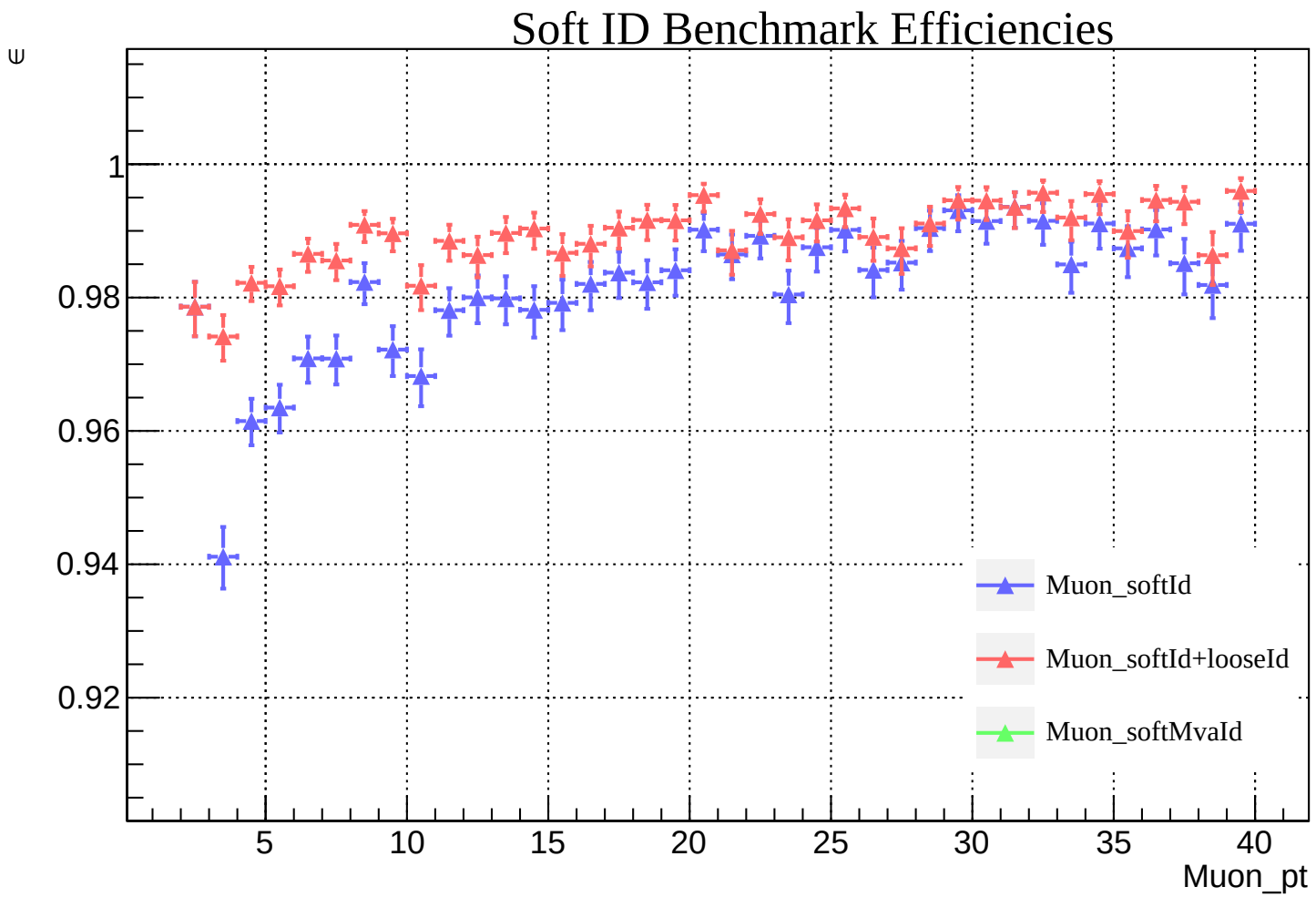
Baseline Model Performances



ϵ = # true muons that pass ID/# true muons

backup

Baseline Model Performances



Project Repository: `https://github.com/Jphsx/KUSoftMVA`