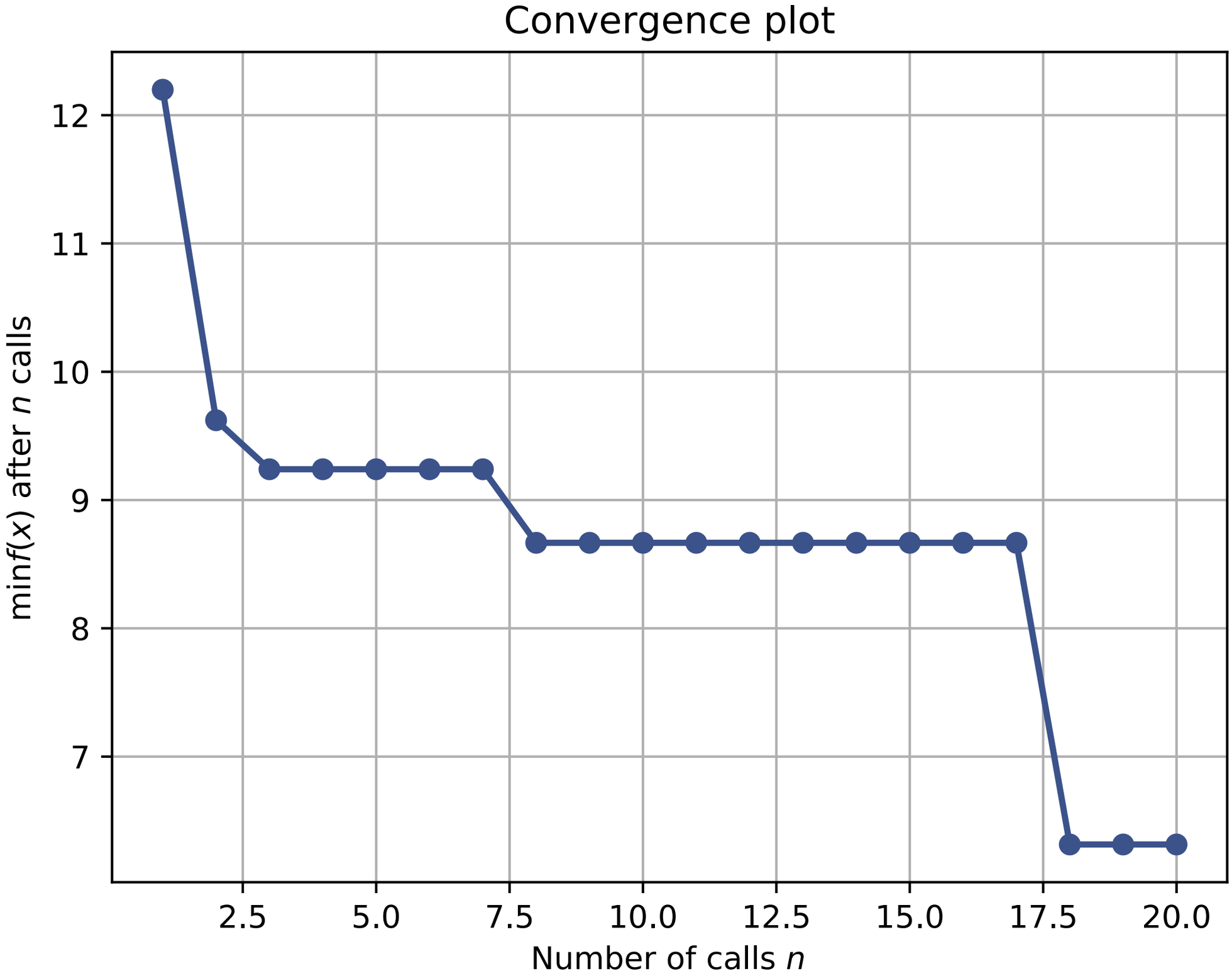
Updates on gamma vs. neutral pion classifier

Wei Wei

University of Illinois at Urbana-Champaign

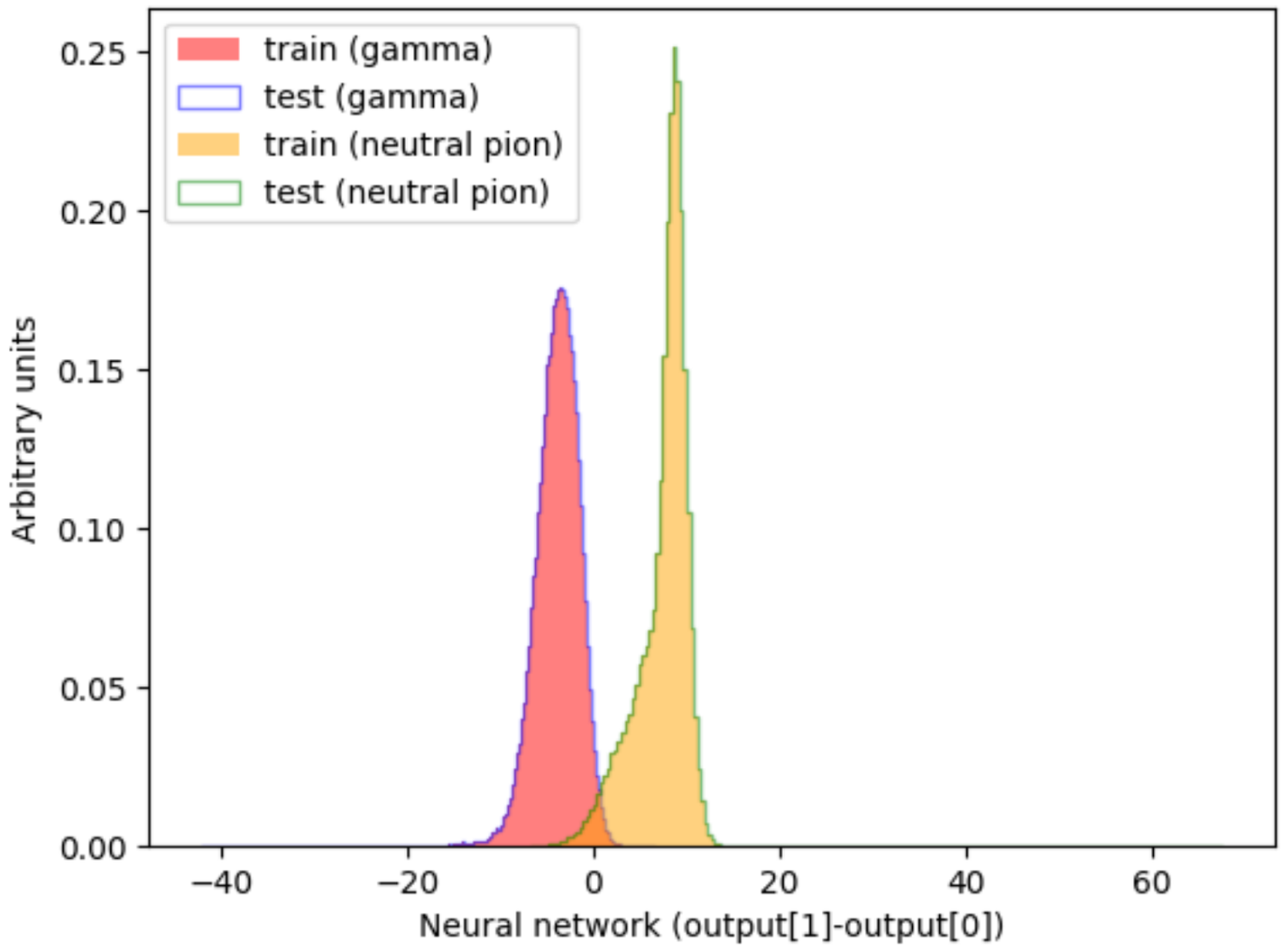
Using skopt module

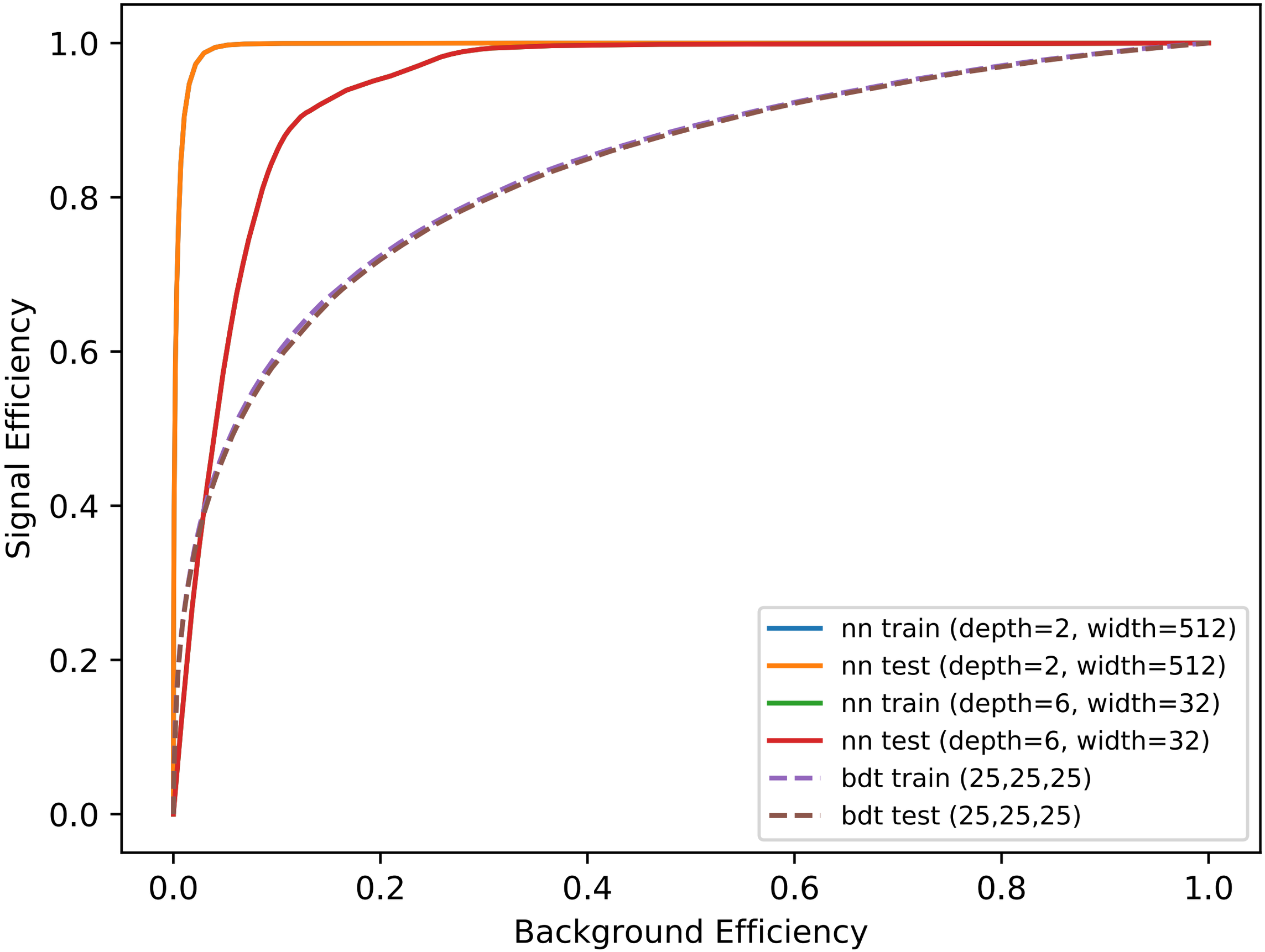
f(x) = (1-acurracy)*100, where x is a point in the space spanned by hyper-parameters



Optimal neural network classifier for gamma and neutral pion: depth=2, width=512

Accuracy > 97%




Signal Efficiency
Background Efficiency
0.0
0.2
0.4
0.6
0.8
1.0
nn train (depth=2, width=512)
nn test (depth=2, width=512)
nn train (depth=6, width=32)
nn test (depth=6, width=32)
bdt train (25,25,25)
bdt test (25,25,25)

About DLKit

Known bugs:

1. When set the variable MaxEvents in Config file to the actual number of events in FileSearch, the generator freezes.
2. The ROC curve returned by the analysis code is broken (not consistent with the output of the neural network)

General problems (Compared to other ML frameworks like PyTorch):

- **User experience:**

1. No documentation. End users have to spend great amount of time on trying to figure out what some function or module does.
2. Not well-tested before released (keep running into bugs). Users should focus on designing better neural network architectures instead of fixing bugs.
3. Not flexible (cannot choose the shape of input or batch size and it's hard to make tmva-like overtraining test plot).
4. The code is long and fragmented, making it hard to trace bugs when they appear.
5. Not actively maintained.

- **Technical aspect:**

1. PyTorch is much faster than Keras on GPU, but DLKit is built on Keras.
2. PyTorch is much more flexible than Keras, which makes it easy to try and build specialized models when going beyond the naive fully connected neural networks.