Contribution ID: **42**   Type: **not specified**

Computing the exact optimal classifier for Ginkgo jets

Tuesday 6 July 2021 16:20 (20 minutes)

In the last several years, the ML4Jets community has worked to improve performance for jet tagging and performed a number of comparisons of different architectures for jet tagging and other tasks. We have seen that combining multiple classifiers together into a meta-tagger or an ensemble improves performance. But is there still room for improvement? In other words, are we approaching the performance of the optimal tagger? Formally, the optimal classifier is defined by a likelihood ratio (Neyman-Pearson lemma), but the likelihood for the observed jet is typically intractable as it involves marginalizing over the enormous number of showering histories. Additionally, the likelihood for a particular shower is, in general, not easily accessible. We consider new datasets with signal and background generated with the Ginkgo model and use the cluster trellis to exactly compute the marginal likelihood under each hypothesis in order to calculate the exact optimal likelihood ratio. As a result, we can compare the performance of ML-based taggers to this optimal classifier.

Affiliation

NYU

Academic Rank

Authors: GREENSPAN, Lauren (NYU); DRNEVICH, Matthew (New York University (US)); MACALUSO, Sebastian (New York University); CRANMER, Kyle Stuart (New York University (US))

Presenter: GREENSPAN, Lauren (NYU)

Session Classification: ML-Assisted Measurements and Searches