10th International Workshop on Boosted Object Phenomenology, Reconstruction and Searches in HEP (BOOST 2018)

Contribution ID: **31** Type: **Talk**

Learning to Classify from Impure Samples with High-Dimensional Data

Tuesday 17 July 2018 15:20 (25 minutes)

Machine learning in high energy physics relies heavily on simulation for fully supervised training. This often results in sub-optimal classification when ultimately applied to (unlabeled) data. In addition to describing a new method for weak supervision (learning directly from data) called Classification Without Labels (CWoLa), we show for the first time how to apply these techniques to high-dimensional data, where significant architectural changes are required. This is critically important for learning from and about the full radiation pattern inside jets.

Authors: NACHMAN, Ben (University of California Berkeley (US)); METODIEV, Eric (Massachusetts Institute of Technology); KOMISKE, Patrick (Massachusetts Institute of Technology); SCHWARTZ, Matthew

Presenters: NACHMAN, Ben (University of California Berkeley (US)); METODIEV, Eric (Massachusetts Institute of Technology); KOMISKE, Patrick (Massachusetts Institute of Technology)

Session Classification: Machine Learning