PHYSTAT - Statistics meets ML



Contribution ID: 31

Type: not specified

Robust signal detection with classifiers decorrelated via optimal transport

Wednesday 11 September 2024 14:50 (25 minutes)

New physics searches are usually done by training a supervised classifier to separate a signal model from the known Standard Model physics (also called the background model). However, even when the signal model is correct, systematic errors in the background model can influence supervised classifiers and might adversely affect the signal detection procedure. To tackle this problem, one approach is to use the (possibly misspecified) classifier only to perform a signal-enrichment step and then to carry out a model-agnostic search in the signal-rich region using only the real experimental data. For this procedure to work, we need a classifier constrained to be decorrelated with one or more protected variables used for the signal detection step. We do this by considering an optimal transport map of the classifier output that makes it independent of the protected variable(s) for the background. We then fit a semi-parametric mixture model to the distribution of the protected variable after making cuts on the transformed classifier to detect the presence of a signal. We compare and contrast this decorrelation method with previous approaches, show that the decorrelation procedure is robust to background misspecification, and analyse the power of the signal detection test.

Primary Field of Research

Presenter: Dr CHAKRAVARTI, Purvasha (UCL) **Session Classification:** Talks