SemiSup - SemiVis

Noam Wunch

December 29, 2020

Contents

1 Introduction

The landscape of machine learning can be broadly separated into supervised, semi-supervised, and unsupervised learning. A promising branch of semi-supervised classification concerns learning from mixed samples. In theory, a model can learn to distinguish signal events from background events given two mixed samples with different signal proportions. This method is often favorable for jet tagging as it allows for weaker model dependence, i.e. the class of models covered by such a search can be wider. A limitation of this method is that it requires the features given to the model to be uncorrelated with the sample labels. In practice this his hard to achieve. If the models of interest have a dijet signature, we can have cuts on one jet separate the events for mixed sample learning using the other jet. The underlying assumption is that some features are largely uncorrelated between the two jets of a given event. We therefore propose a mixed-samples classification method applicable to BSM models that:

1. Have some known feature that can weakly separate SM jets from BSM jets.

2. Have a dijet signature.

2 Learning from mixed samples

3 Co-training

4 Classification method

Given a BSM model with a dijet signature, an event collection $\mathcal{D}$, and a weak jet classifier $h^{(0)}$:

1. Predict the labels $L_1^{(0)}$ and $L_2^{(0)}$, corresponding to the predictions of $h^{(0)}$ on the first and second jets respectively.

2. for $1 < i \leq N_{iter}$

    (a) Train a classifier $h_1^{(i)}$ using the labels $L_2^{(i-1)}$.
    (b) Train a classifier $h_2^{(i)}$ using the labels $L_1^{(i-1)}$.
    (c) Predict the labels $L_1^{(i)}$ and $L_2^{(i)}$, corresponding to the predictions of $h_1^{(i)}$ and $h_2^{(i)}$ respectively.

3. Combine $L_1^{(N_{iter})}$ and $L_2^{(N_{iter})}$ for a final event classification L.

All training and prediction steps are applied to all examples in $\mathcal{D}$.

5   Test on benchmark model

A   Code implementation

The code can be fetched from github via

$$\texttt{git clone https://github.com/noamwunch/SemiSupSemiVis}$$

And executed with `./RUN/run_example.sh`.

References