CS281 Status Update

Scalable Signal Region Identification with applications to the ATLAS $W^{\pm}W^{\pm}W^{\pm}$ Analysis

Nicolo' Foppiani, Jonah Philion, Baojia(Tony) Tong

Abstract

We want to find signal regions and control regions blah blah blah. We use machine learning to discover physics efficiently.

Problem Statement In particle physics analysis, there is in general some "signal"process that generates data simultaneously as "background"processes generate similar looking data. Given unlabeled data from all processes, our goal is to determine the likelihood that the signal process exists.
To generate data $\{x_i\}_N$ lying in $\mathbb{R}^n$, we associate to the signal process a smooth function $s : \mathbb{R}^n \to \mathbb{R}_{>0}$ and to the background process $b : \mathbb{R}^n \to \mathbb{R}_{>0}$ such that for any volume $V \subset \mathbb{R}^n$, the number of background events in this region $n_{bkg}(V)$ and the number of signal events in this region $n_{sig}(V)$ can be modeled as

$$n_{bkg}(V) \sim \text{Pois}\left(\int_V b\right)$$

$$n_{sig}(V) \sim \text{Pois}\left(\int_V s\right)$$

The data is generated from these Poisson processes. We can then determine the likelihood of the existence of the signal process by choosing regions of our parameter space $\{V_j\}$ then comparing the number of events observed in these regions to the values $\{n_{bkg}(V_i)\}$ and $\{n_{sig}(V_i) + n_{bkg}(V_i)\}$. For the case of large N, we can approximate the Poisson distribution as Gaussian. For a region with O observed events, the significance is then

$$Significance \approx \frac{O - n_{bkg}(V)}{\sqrt{n_{bkg}(V)}} \tag{1}$$

We have simplified our problem to determining the set $\{V_i\}$ of regions of parameter space which will maximize the definition of significance in Equation 1.

Baselines Since our goal is to determine volumes of parameter space which maximize the significance, one baseline is to take the volume to be the entire paramter space. The metrics for this approach are summarized in Figure 1.

In particle physics, we need significance greater than 3σ to publish, so the fact that no cuts at all produces a significance of 0.517σ means that we need to use smarter approaches to discover WWW decay.

A description of the data used in this analysis is in Figure 2.

	n_{sig}	n_{bkg}	$n_{sig}/\sqrt{n_{bkg}}$
No Cuts	47.09	8288.18	0.517σ

Fig. 1: Baseline measurements. n_{sig} is the weighted sum of all signal events, n_{bkg} is the weighted sum of all background events, and $S/\sqrt{B}$ is our metric for significance. We would like to maximize the significance to some value above 3 in particle physics.

Variable Name	Description
$j0_m$	something cool
$j0_{pt}$	something cool
$j0_{eta}$	something cool
$j0_{phi}$	something cool
$l0_m$	something cool
$l0_{pt}$	something cool
$l0_{eta}$	something cool
$l0_{phi}$	something cool
$l0_c$	something cool
$l0_{isEl}$	something cool
$l1_m$	something cool
$l1_{pt}$	something cool
$l1_{eta}$	something cool
$l1_{phi}$	something cool
$l1_c$	something cool
$l1_{isEl}$	something cool
$l2_m$	something cool
$l2_{pt}$	something cool
$l2_{eta}$	something cool
$l2_{phi}$	something cool
$l2_c$	something cool
$l2_{isEl}$	something cool
met_{pt}	something cool
met_{phi}	something cool
$weight$	something cool
cl	The process that generated event. In our analysis, we only include WWW and WZ.
is_{sig}	1 if the process which generated this event is signal. 0 otherwise.

Fig. 2: Feature Space. We measure cl, is_{sig}, and $weight$ only in simulation. We measure all other variables in real data and in simulation. Our general approach to train classifiers to predict is_{sig} given the jet and lepton features.

Approaches

1. **Naive Classification (Jonah)** In which I train $x \to y$ classifier where x is a list of kinematic and categorical variables and y is 0 for background and 1 for signal.
2. **Category Specific Classification (Jonah)** In which I group the data by the categorical variables, then train a classifier on each of these groups (I get above 2 sigma with this).
3. **Soft Significance (Jonah)** I make the significance differentiable then do gradient descent on it. I haven't gotten this to work better than the No-Cut baseline.
4. **Neural Network (Nico)** Fun with PyTorch
5. **Physics Motivated Cuts (Tony)** Cuts from the most recent analysis paper