SemiSup - SemiVis Documentation

Noam Wunch

February 3, 2021

Contents

1 Pseudo-search

Our proposed search is sensitive to $Z' \to X\bar{X}$ signals with higher-than-QCD jet constituent multiplicity. The search should be stratified in $m_{Z'}$. Candidate dijet events are selected according to $m_{Z'}$ region and other standard cuts. The N events that passed the cuts are split randomly into k groups. One of the groups is pulled aside as a test set and the following steps are applied to the remaining events:

1. Obtain two sets of weak labels - L_1 and L_2. One by cutting on jet_1 multiplicity (L_1) and the other by cutting on jet_2 multiplicity (L_2). Where jet_1 and jet_2 are the leading PT and next-to-leading PT jets in the event, respectively.

2a. Train a classification model on jet_1 data, using sig% of most signal like events and bkg% of most background like events, to obtain the classifier h_1.

2b. Train a classification model on jet_2 data, using sig% of most signal like events and bkg% of most background like events, to obtain the classifier h_2.

3. A final event classifier, h, is the arithmetic mean of h_1 and h_2.

Apply h to the test set to obtain a (stronger) prediction of how signal like each event is. Repeat items 1-3 k times, each time using a different group as test set. Once all events have a prediction, cut on the predictions to obtain a more signal rich event collection. Using this event collection, conduct a bump-hunt in M_{jj}, looking for excess events at M_{jj} near $m_{Z'}$. Repeat for other regions of $m_{z'}$ (while keeping track of look-elsewhere effect).

2 Benchmark dataset

We applied our pseudo-search to $N = 150k$ events that passed event selection, with signal fraction $s = 5\%$. Our benchmark signals are pair-created, displaced, semi-visible, dark jets. We considered hidden-sector models with invisible fraction $r_{inv} = \{0, 0.25\}$ and dark-pion decay length of 0.5 mm. For this choice of displacement the leading background contribution is SM $b\bar{b}$ events. The events were generated using Pythia's Hidden-Valley model. The exact signal parameters are listed in appendix A. In all cases, $m_{Z'} = 1$ TeV.

2.1 Event selection

We selected for the following dijet events:

1. $M_{jj} > 500$ GeV
2. jet-PT > 50 GeV (both jets)
3. |jet-rapidity| < 2.5
4. b-tagging? displacement?

2.2 Multiplicity

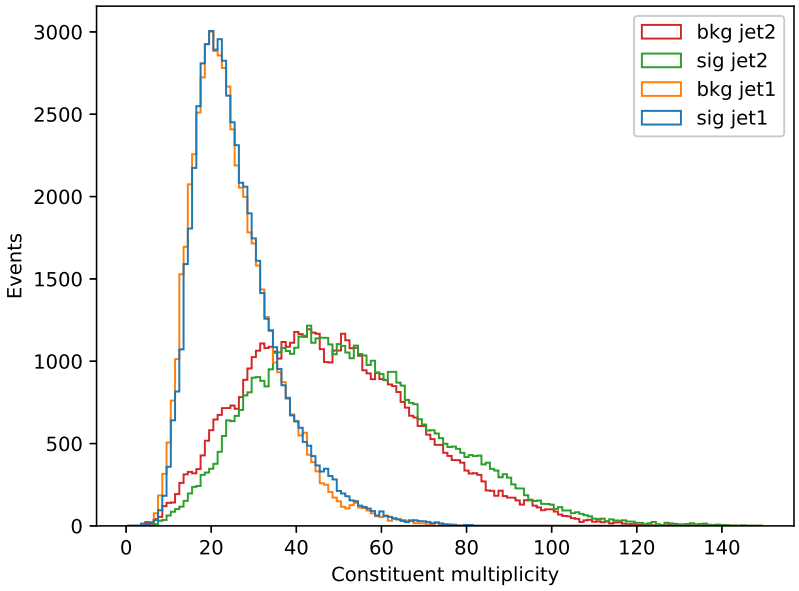

Figure 1: Multiplicity

2.3 MET

2.4 Leading dark jets

2.5 $t\bar{t}$ contribution

3 Jet classification

We chose an LSTM network as our jet classifier. Each event is represented by its two, P_T-wise, leading jets. Tracks and towers are aggregated using Particle-Flow. Jets are reconstructed from Particle-Flow constituents using k_t nearest neighbour algorithm with $\Delta R = 0.7$.

3.1 Jet representation

Each jet is represented by a list of constituents, ordered by P_T. Constituent information consists of the 5D perigee parameters (p_T, η, ϕ, d_0, and d_z) of each track in the jet, together with angular distance from jet axis ΔR. The track information was transformed to the parton frame and p_T was scaled by jet-p_T (p_T^{jet}), i.e.:

$$\bar{p}_T = p_T/p_T^{\text{jet}}\,, \quad \Delta\eta = \eta - \eta^{\text{jet}}\,, \quad \Delta\phi = \phi - \phi^{\text{jet}}\,, \quad \bar{d_z} = d_z/\cosh(\eta^{\text{jet}}), \quad d_0, \quad \text{and} \quad \Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}.$$

For neutral particle flow constituents displacement parameters, d_0 and d_z, were set to zero.

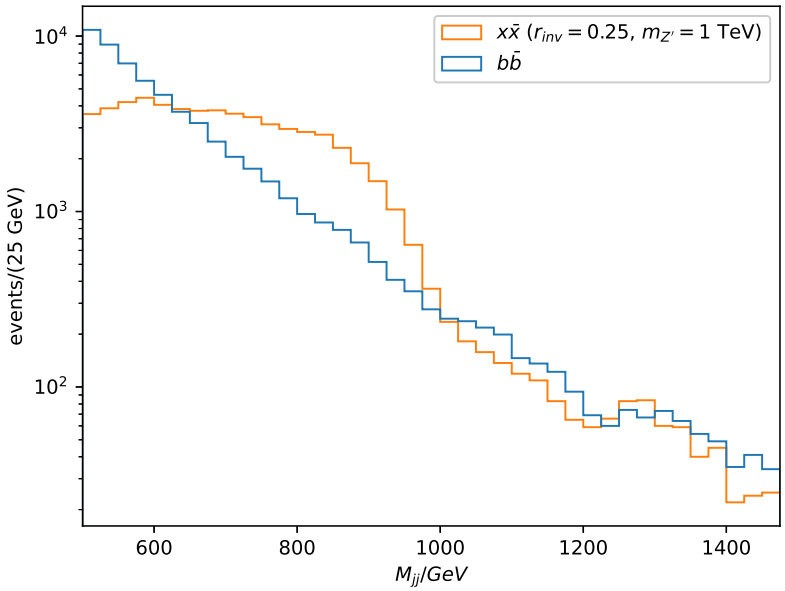

Figure 2: M_{jj}

3.2 Model architecture

We used an LSTM classifier with three layers. The first layer consisted of 50 LSTM cells, the second was a dense layer with 16 nodes and the last was dense layer with one node. The network requires the constituent lists to be of fixed length. Therefore, the track lists were truncated/masked to a fixed length, $N_{constits}$, which we set to 80. The NN input of a single jet was thus a $(N_{constits} \times N_{features})$ matrix.

4 Results

A Benchmark signal parameters

We used a hidden sector with the following parameters

- $m_{Z'} = 1$ TeV
- ProbV $= 0.75$
- dark pions decay length $= 0.5$ mm

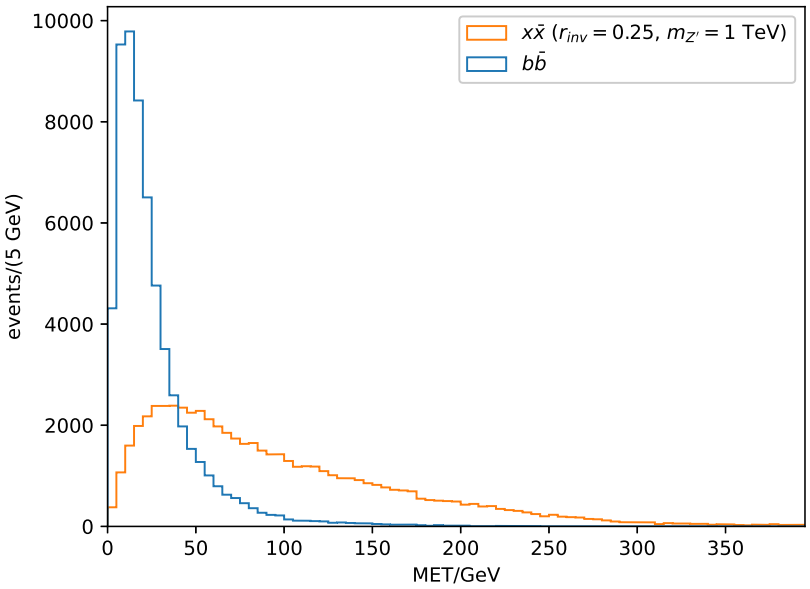
$x\bar{x}$ ($r_{inv} = 0.25$, $m_{Z'} = 1$ TeV)
$b\bar{b}$
events/(5 GeV)
MET/GeV
0
2000
4000
6000
8000
10000
0
50
100
150
200
250
300
350

Figure 3: MET