Top background estimation

Fredrik Tegenfeldt (ISU)
Jim Cochran (ISU)
Jan Conrad (CERN)

- Introduction
- Background estimation from data
- Multivariate methods
- New method: RuleFit
- SUSY vs Top using RuleFit
- Conclusions
Introduction

- SUSY signature: multi jet and large missing $E_T$
- Backgrounds:
  - Top events
  - QCD tails
  - $W$ plus jets
- How to separate background and signal?
  - Estimate background from data
  - Multivariate methods
Top background from data

- Previously explored by Dan Tovey (Software week May 25, 2005)
- The method proposed
  - Reconstruct top mass from hard jet and lepton + missing $E_T$ with $W$ mass constraint
  - Estimate background using top mass signal and side band
- Drawbacks
  - Method relies on good missing $E_T$ and jet $E_T$ resolutions
  - Does not use b-tagging
  - Tevatron experience: top mass estimator biased depending on measuring technique
- Idea: look at variables not depending on missing $E_T$ and only weakly on jet $E_T$ resolutions
From Tevatron top mass measurement:

- Aplanarity vs. $H_T$ - was used to control W+jets

Doesn't look that attractive
Multivariate Methods

- Several methods exist:
  - Sequential cuts
  - Likelihood ratios
  - Fisher discriminants
  - Neural Networks
  - Decision trees

- New method: **RuleFit** (J.H. Friedman, ref next page)
  - Very efficient
  - Relatively fast
  - “Grey” box
  - Implementation for \textbf{R} (\url{www.r-project.org/})
RuleFit

- Predictive model (simplest case):

\[ F(x) = \hat{a}_0 + \sum_{m=1}^{M} \hat{a}_m r_m(x) \]

- Examples of rules

\[ r_1(x) = I(M_{\text{eff}} > 100.0) \]
\[ r_2(x) = I(M_{\text{eff}} \leq 100.0) \cdot I(p_T^j < 200) \]

- Coefficients from minimizing a loss-function \( L(y, F(x)) \)
- Rule ensemble generated using decision trees
- Details – see http://www-stat.stanford.edu/~jhf/R-RuleFit.html
RuleFit – the input

- Training samples; signal and background
- Some tuning related to:
  - the minimization
  - decision trees

Very little tuning is needed – it worked fine out of the box
RuleFit – the output

- The predictive model (coefficients and rules)
- A measure of importance of a variable
  - Function of:
    - How often they appear in predictors (rules)
    - The absolute value of the coefficients

**Useful to exclude unimportant variables**

- Some others not yet explored by us

- For the data, the score function $y = F(x)$
  - signal like : high values
  - background : low values

This is what you cut on
SUSY and RuleFit

- Signal samples:
  - SU1 Coannihilation
  - SU3 Bulk region
- Background: Top (T1)
- Training samples
  - SU1 : 10000
  - Top (T1) : 10000
- Signal samples
  - SU1, SU3 : 75000 events each
### Obtained rules – the 4 most important

<table>
<thead>
<tr>
<th>Rule</th>
<th>Condition</th>
<th>Coefficient</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$246.8 &lt; E_T &lt; 2180$</td>
<td>0.1530</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>$20.45 &lt; p_T^{jet} &lt; 236.2$, $E_T &lt; 145.5$</td>
<td>-0.1371</td>
<td>96.11</td>
</tr>
<tr>
<td>3</td>
<td>$142.5 &lt; M_{eff} &lt; 738.6$, $E_T &lt; 205.3$</td>
<td>-0.1369</td>
<td>95.78</td>
</tr>
<tr>
<td>4</td>
<td>$139.4 &lt; M_{eff} &lt; 681.5$, $E_T &lt; 193.5$</td>
<td>-0.1215</td>
<td>85.17</td>
</tr>
</tbody>
</table>
Relative Importance of Input Variables

- $M_{\text{eff}}$
- $p_{\text{jet}}^{T,1}$
- $p_{\text{jet}}^{T,2}$
- $H_T$
- $H_{jj}$
- Sphericity
- $N_{\text{elec}}$
- Aplanarity
- $N_{\text{bjets}}$
- $H_E$
- $N_{\text{muon}}$
- Planarity

Variable Importance

Input Variable
Scores on a SU1 trained model

Top T1
Susy SU1

Not scaled to x-sections!
Signal and background efficiencies

SU1 using SU1 trained model
SU3 using SU1 trained model

Background is ttbar
Conclusions and outlook

- Using a top mass estimator for background estimation seems difficult – at least initially
- Aplanarity vs $H_T$ does not look very promising
- RuleFit looks like a promising tool
  - once you have the model, it's easily applied to data
  - variable importance is useful
- Next we want to incorporate $W+$jets background
- What about negative event weights in top?
  - Not obvious what that means when used in a learning machine
  - Check scores for negative and positive weights separately
- Further explore RuleFit in SUSY background context
EXTRA SLIDES
Discriminators

List derived from published top analyses at CDF and D0

- Missing $E_T$
- Jet $p_T$ or sum of jet $p_T$
- $H_T$ excl. 2 hardest jets
- Sum of jet $E$
- Centrality
- Effective mass
- Aplanarity: $\sim$smallest eigenvalue of the momentum tensor