Gradient Boosted Decision Trees and Particle Physics

Aaron Webb

October 27, 2017

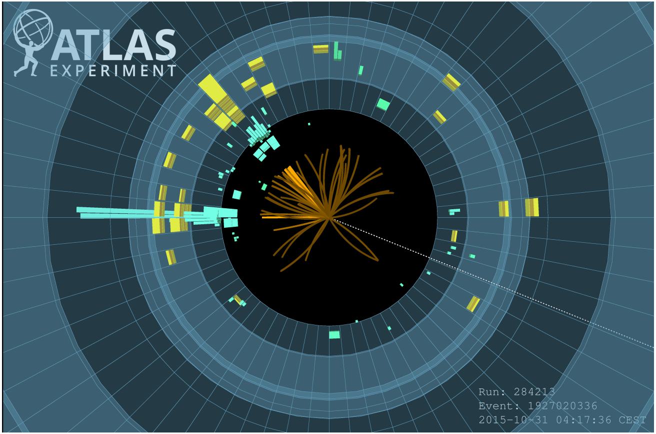
The LHC and Big Data

- Bunches of 10^{11} protons are collided every 25 ns
- Produces ≈ 50 PB of data per year
- Particle lifetimes $\mathcal{O}(10^{-25})$ seconds, only ever see decay products

- Many processes look the same in the detector
- Interesting interactions are rare

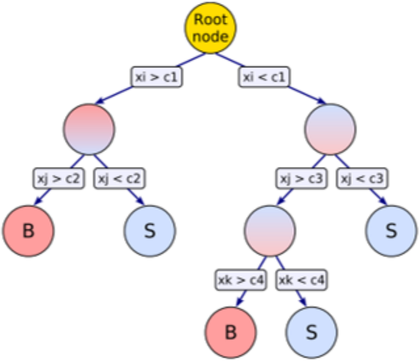
The ATLAS Dectector

- Tells us the types of particles, their momentum, energy, and location
- Use these to predict what happened



Gradient Boosted Decision Trees

- Combines a set of weak "learners" into a single "strong" learner
- Start with a simple model - single binary decision tree
- Construct a new tree to correct the weaknesses of the model
- Iterate till it converges



The Algorithm

- Create a model, $F(x)$, for a data set y
- Start with a simple model, $F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$
- Model the residuals: $h(x) = y - F_0(x)$
- Update the model: $F_{m+1}(x) = F_m(x) + h(x)$
- Iterate till the model converges
- $F(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + \text{const}$

Improvements

- Shrinkage
  - $F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m h_m(x), \quad 0 < \nu \leq 1$
  - ν is the "learning rate", typically <0.1
  - Improves results, but increases computation costs
- Stochastic Boosting
  - Each successive tree is fit to a random subsample
  - Prevents overfitting, improves speed
- l_2 penalty term
  - Penalize complex trees, remove branches that produce no differentiation

Pros and Cons

Pros

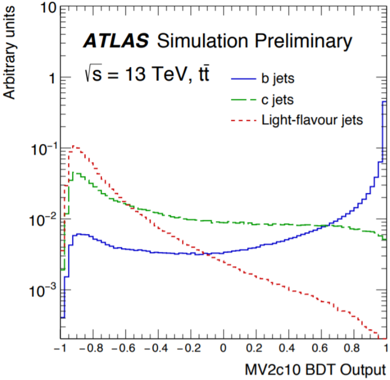
- Easy to use once model has been developed
- Few input parameters needed to tune
- General framework, relevant for a large number of applications

Cons

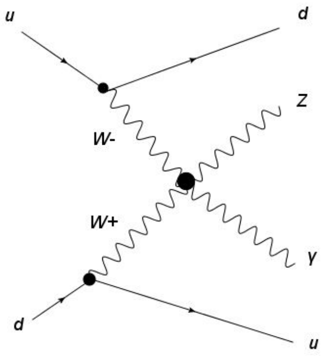
- Training the model can be slow
- Difficult to interpret the output
- Not ideal for sparse data, large numbers of features

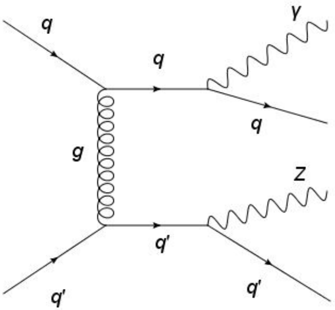
Uses in Particle Physics

- Separate signal and background events
  - Use Monte Carlo simulations to train the model, use for data
- Distinguish ”real” particles from ”fakes”
  - Particles misidentified by the detector, from secondary sources
- ”B-tagging” - identifying different flavors of quarks
  - Quarks ”hadronize”, leaving complex signatures



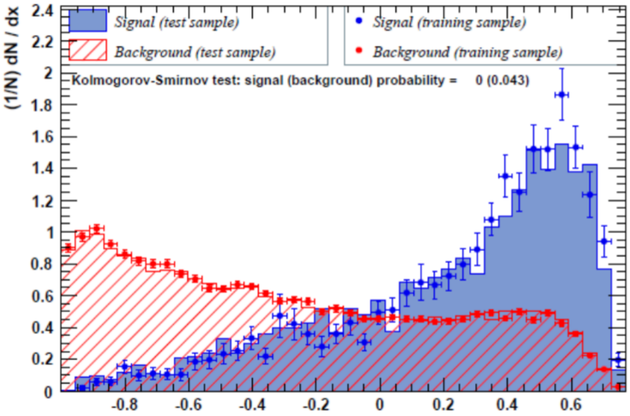
Vector Boson Scattering



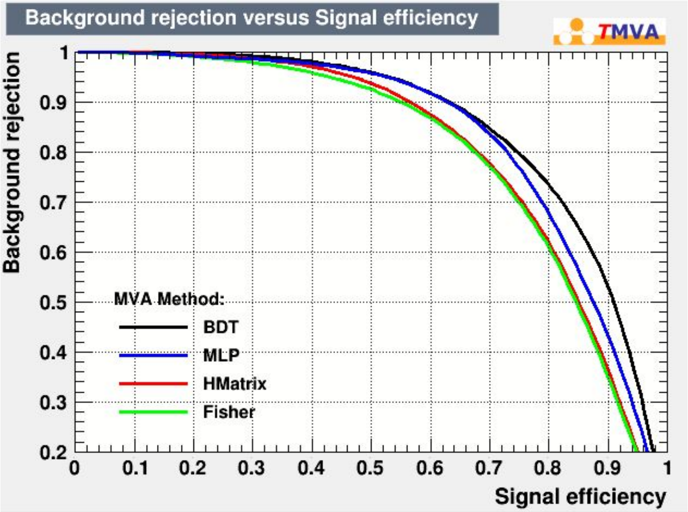


Results

Input Variables	
M_{jj}	η_γ
ΔY_{jj}	$\Delta R_{\mu\gamma}$
$\Delta R_{j\gamma}$	$\Delta R_{\mu\mu}$
η_μ	$p_{T\gamma}$
p_{Tjet}	$p_{T\mu}$



Results



Background rejection versus Signal efficiency
TMVA
Background rejection
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Signal efficiency
MVA Method:
BDT
MLP
HMatrix
Fisher