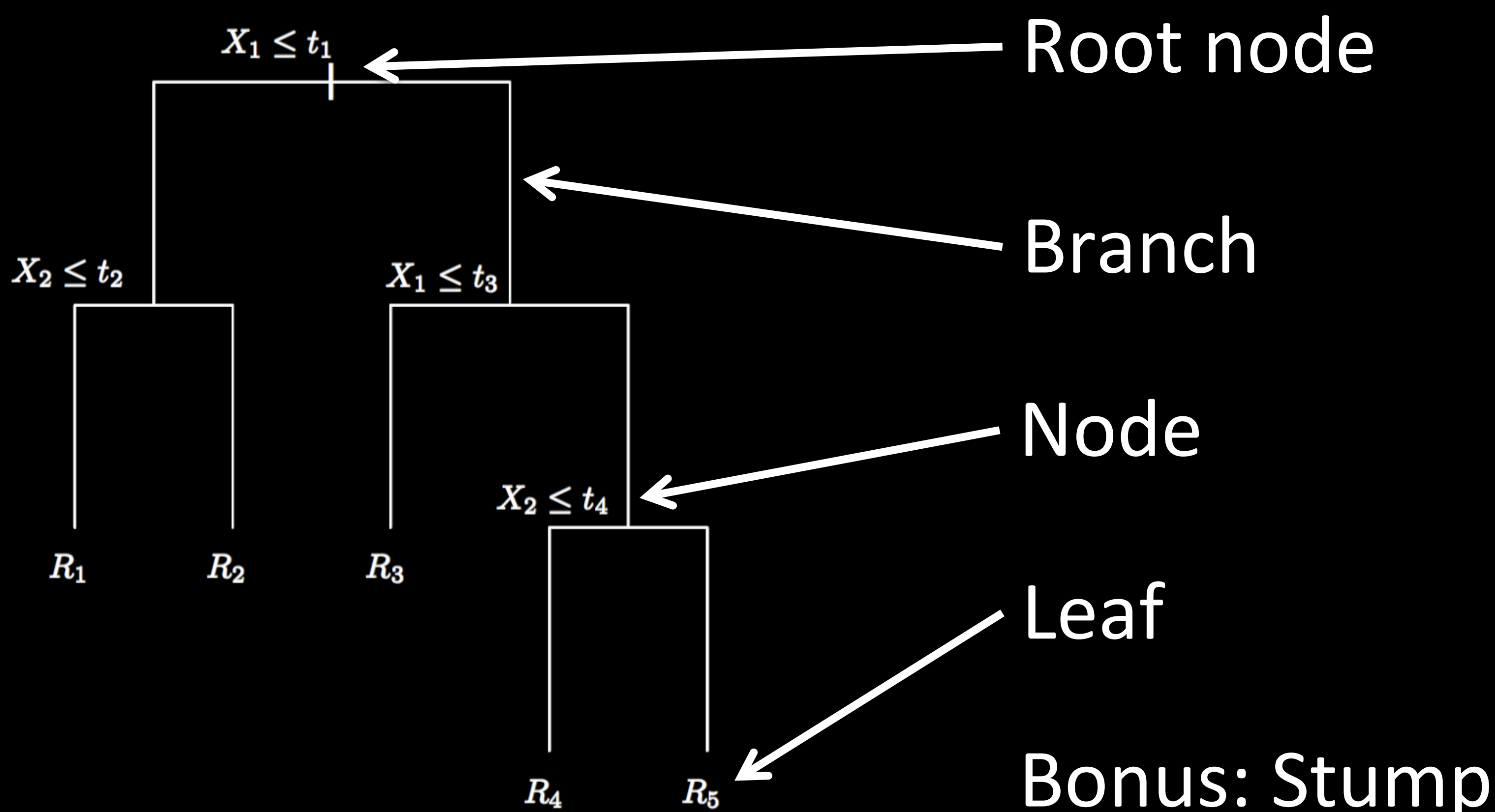
b-Tagging with Boosted Decision Trees

Andreas Biekert

Physics 290E

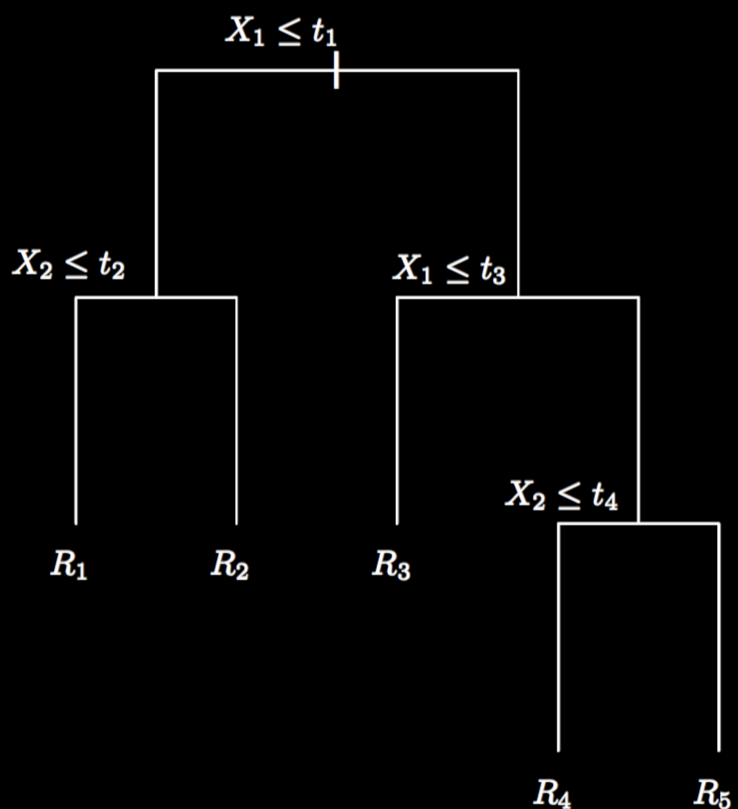
September 20th, 2017

Decision Tree Basics



G. James, et al. *An Introduction to Statistical Learning*.

Decision Tree Basics



G. James, et al. *An Introduction to Statistical Learning*.

Decision Tree Details

- Growing process
  - Top-down: begin with whole data set; split once
  - Recursive: repeat on two new data sets
  - Greedy: make best split at the current step; don't consider future trees
- Decision-making parameter:

$$E = 1 - \max_k(\hat{p}_{mk})$$

Classification Error

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk}$$

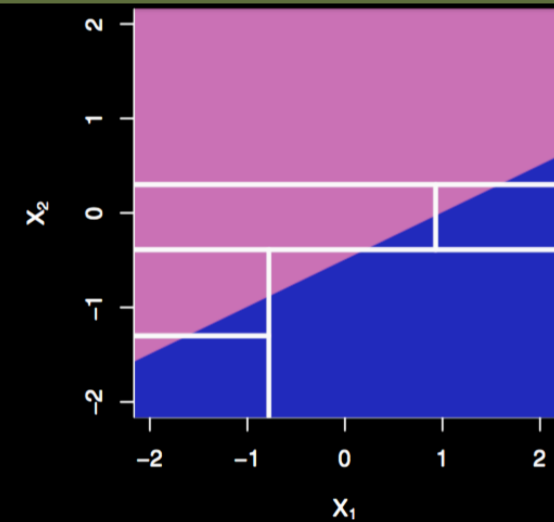
Entropy

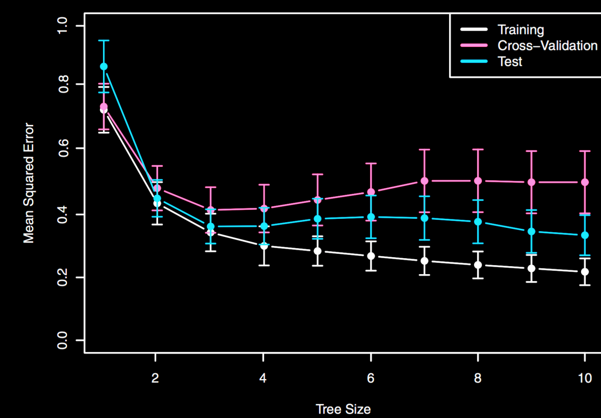
$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

Gini Index

Decision Tree Details

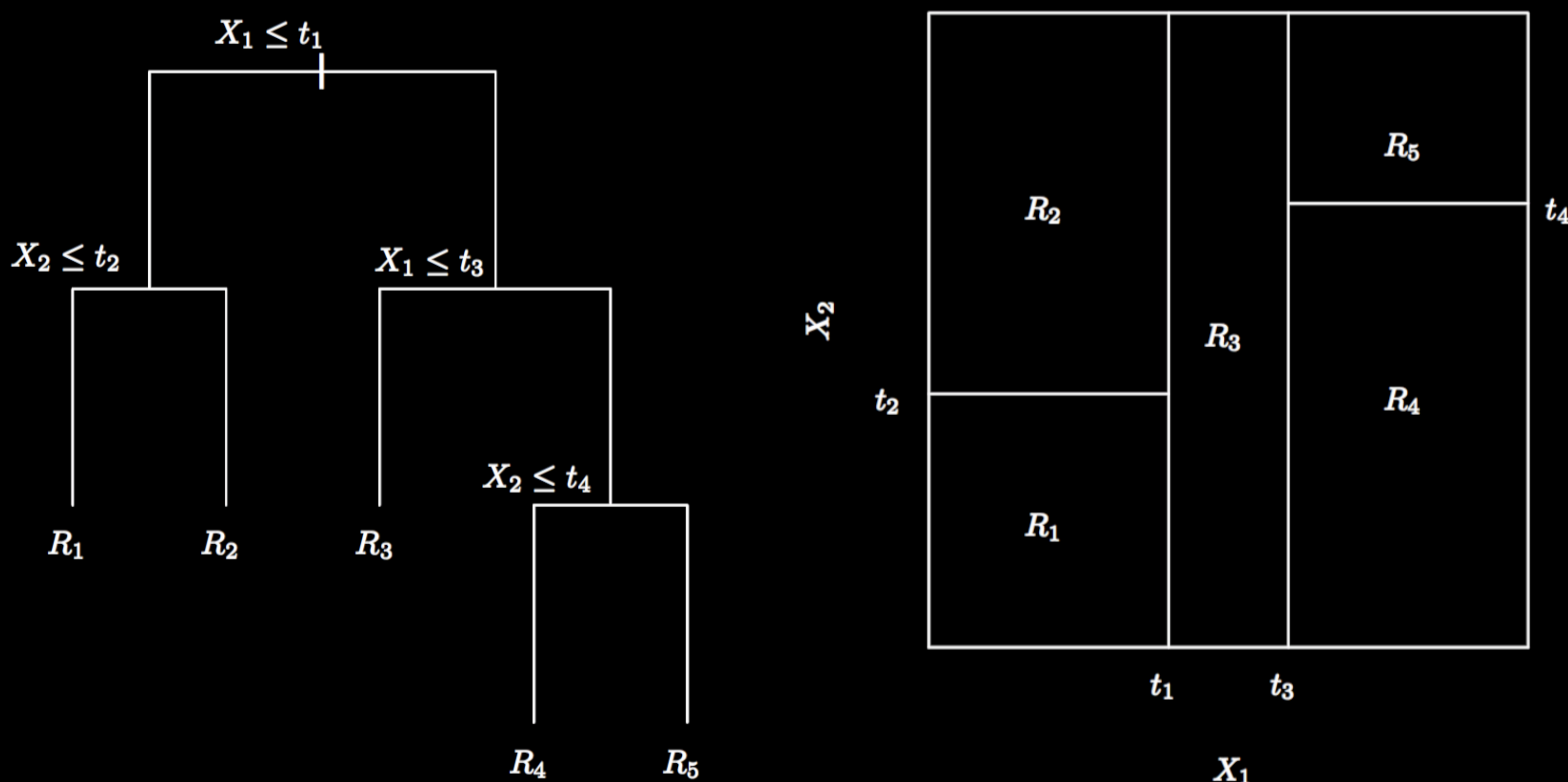
- Overfitting
  - Growing a tree to arbitrary depth can pick up spurious features
- Pruning
  - Cost complexity function: $R_\alpha(T) = R(T) + \alpha N_T$
  - Prune worst subtree at each node and generate list of subtrees. Pick one through validation sample





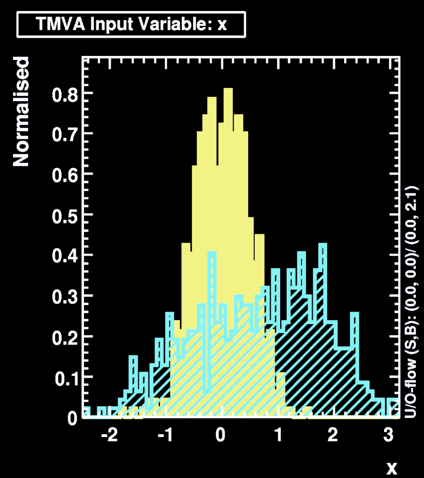
G. James, et al. *An Introduction to Statistical Learning*.

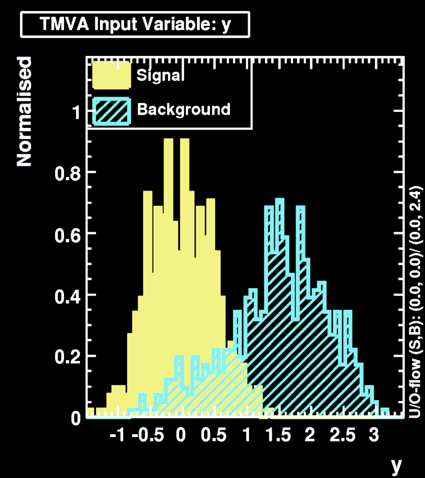
Decision Trees

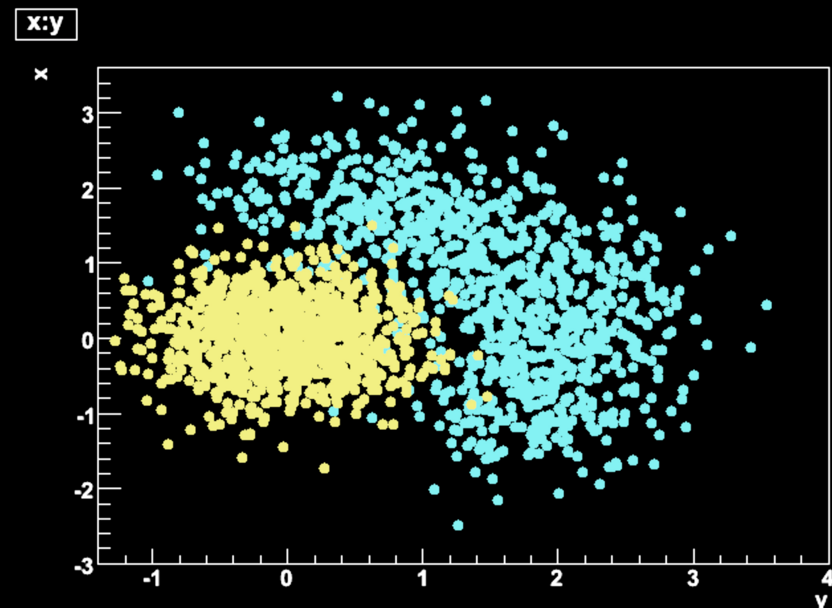


G. James, et al. *An Introduction to Statistical Learning*.

Decision Tree Walkthrough

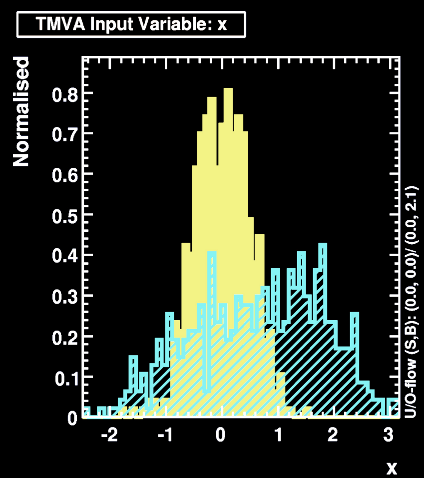


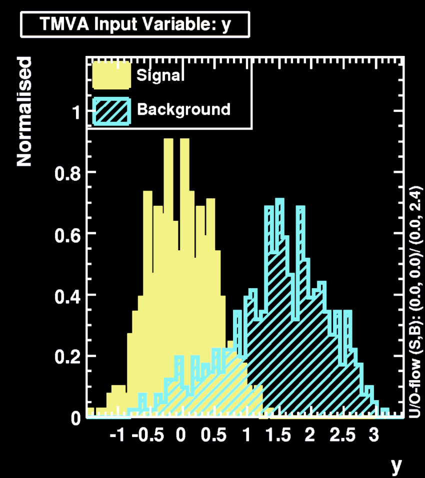


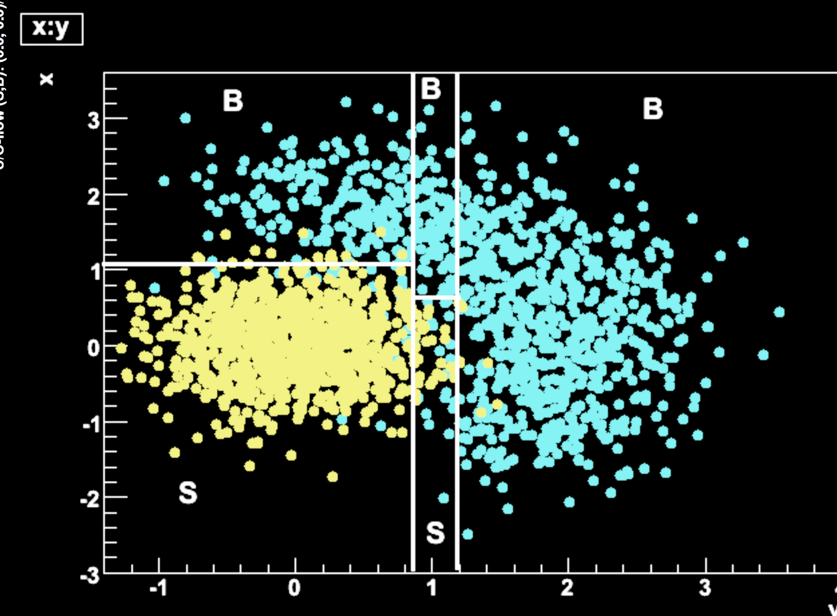


Y. Coadou. European School of Instrumentation in Particle and Astroparticle Physics, Feb. 2016.

Decision Tree Walkthrough

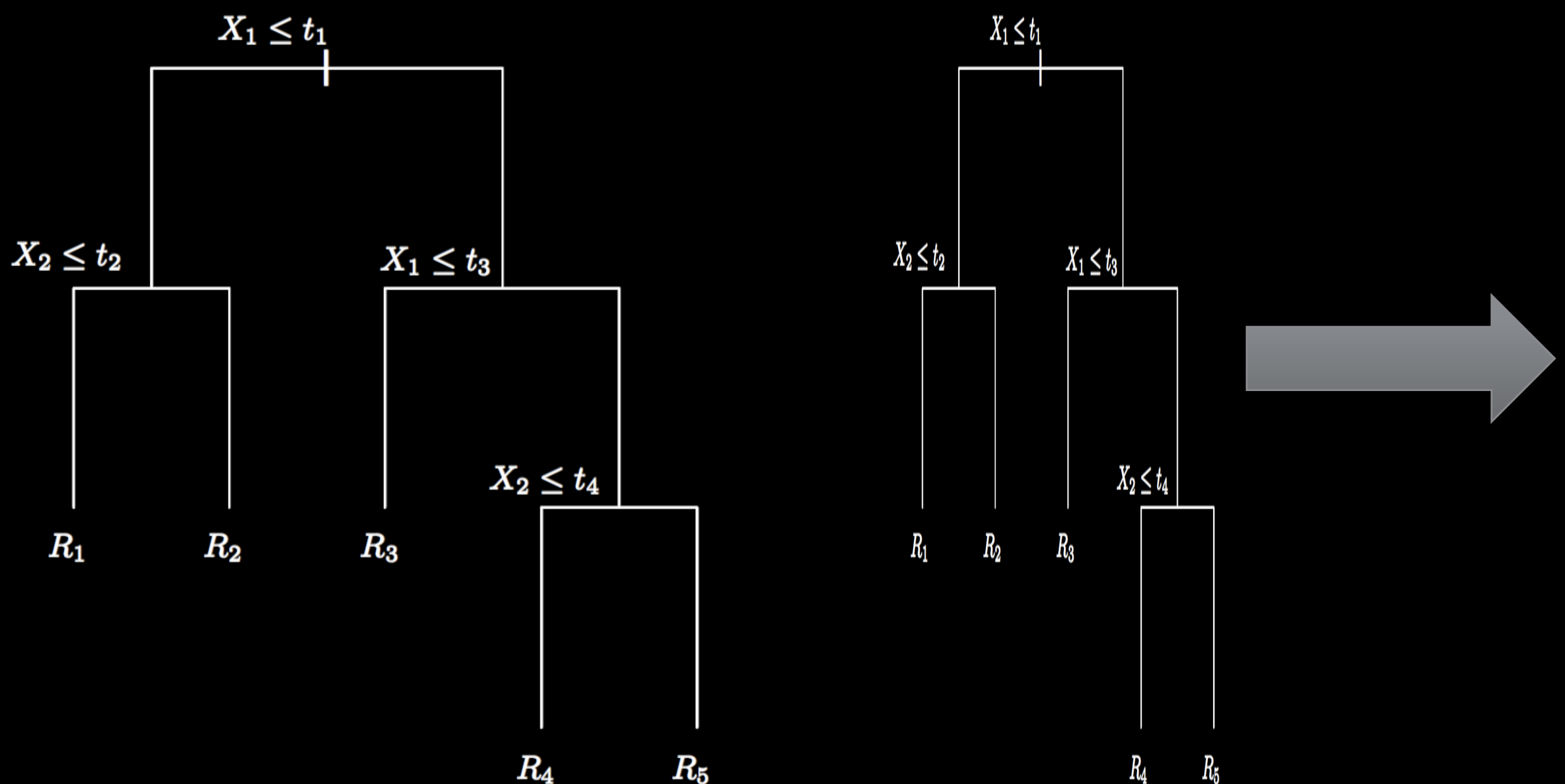






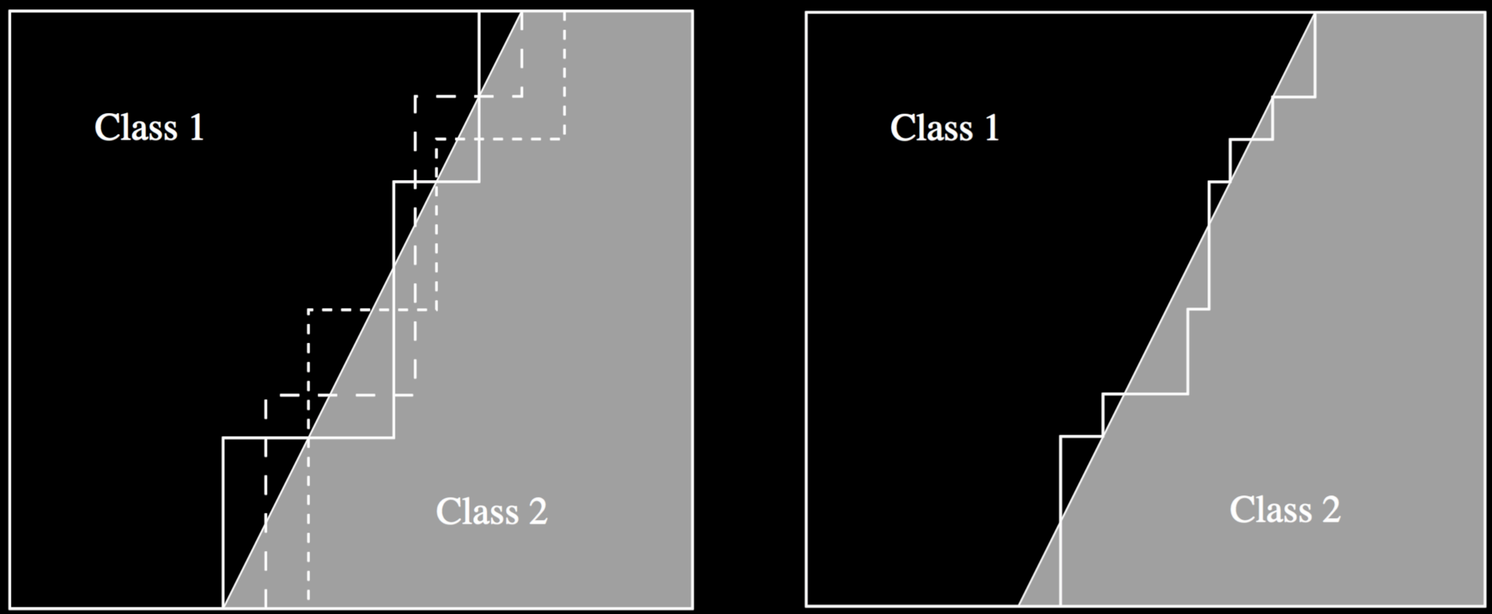
Y. Coadou. European School of Instrumentation in Particle and Astroparticle Physics, Feb. 2016.

Boosted Decision Trees (BDTs)



G. James, et al. *An Introduction to Statistical Learning*.

Ensemble Learning



- Have multiple trees vote on final output.
  - Need a way to produce varied trees.
  - Random forests; bagging

T Diettrich. AI Magazine, 18(4): 97–136, 1997.

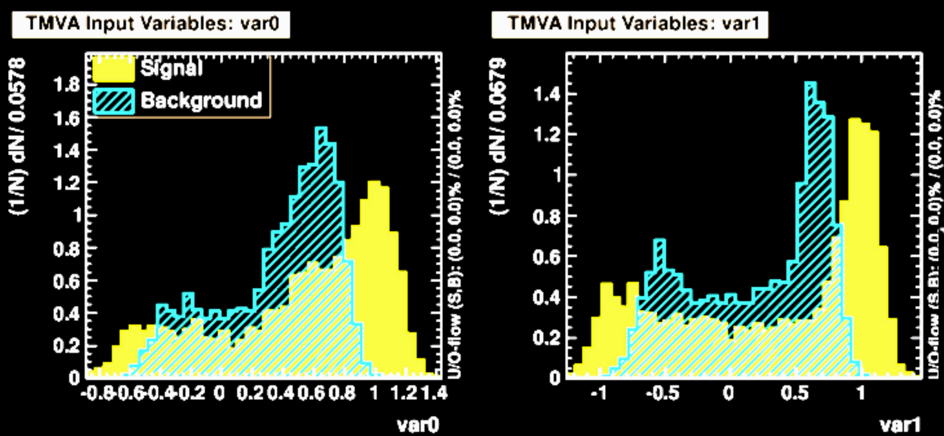
Boosting

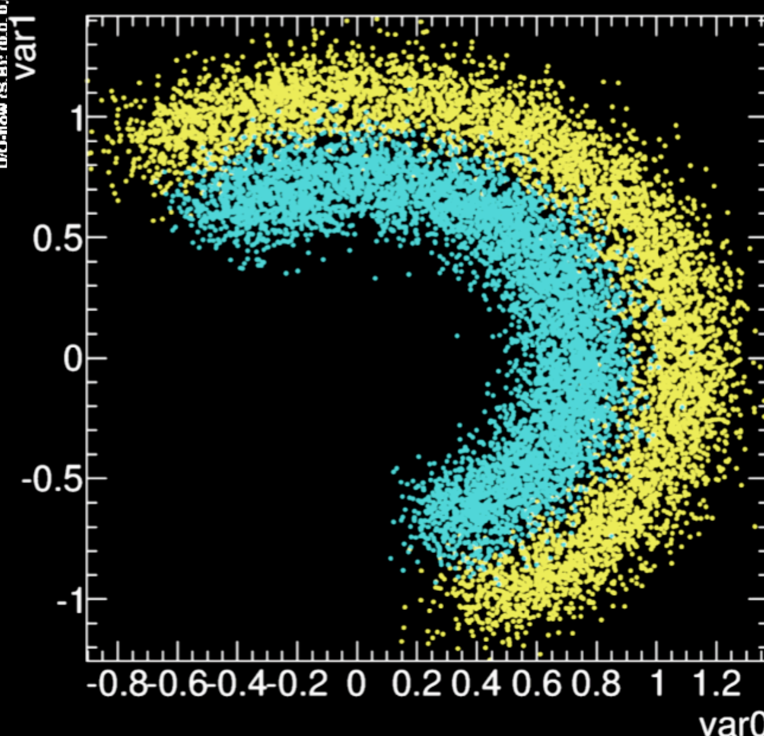
- The logic is to increase weight of misclassified data after building a tree.
  - For example, AdaBoost algorithm has weights w_i^k for the *i*-th datapoint in the *k*-th tree.
  - After each tree, update weights by

$$w_i^k \to w_i^{k+1} = w_i^k \times e^{\alpha_k}, \;\; \alpha_k = \beta \times \ln((1-\varepsilon_k)/\varepsilon_k)$$

  where ε_k is the misclassification rate and β scales tree weights α_k.
  - Boosted result: $T(i) = \frac{1}{\sum_{k=1}^{N_{\mathrm{tree}}} \alpha_k} \sum_{k=1}^{N_{\mathrm{tree}}} \alpha_k T_k(i)$

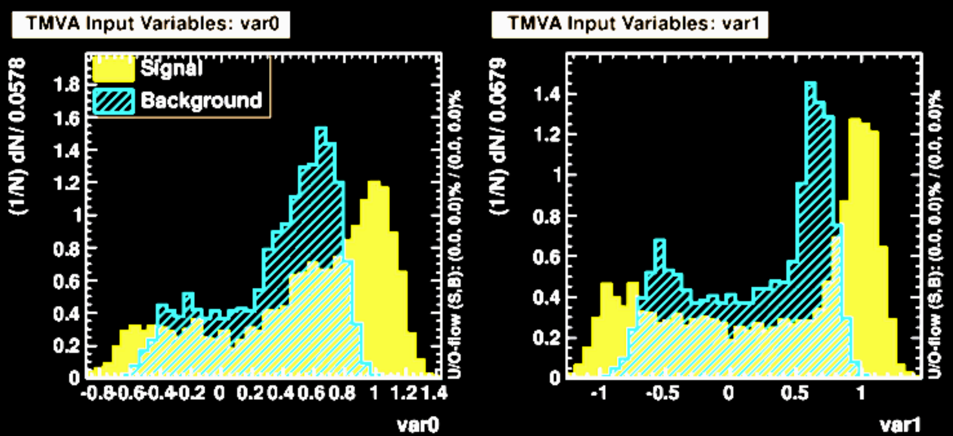
Boosting Walkthrough

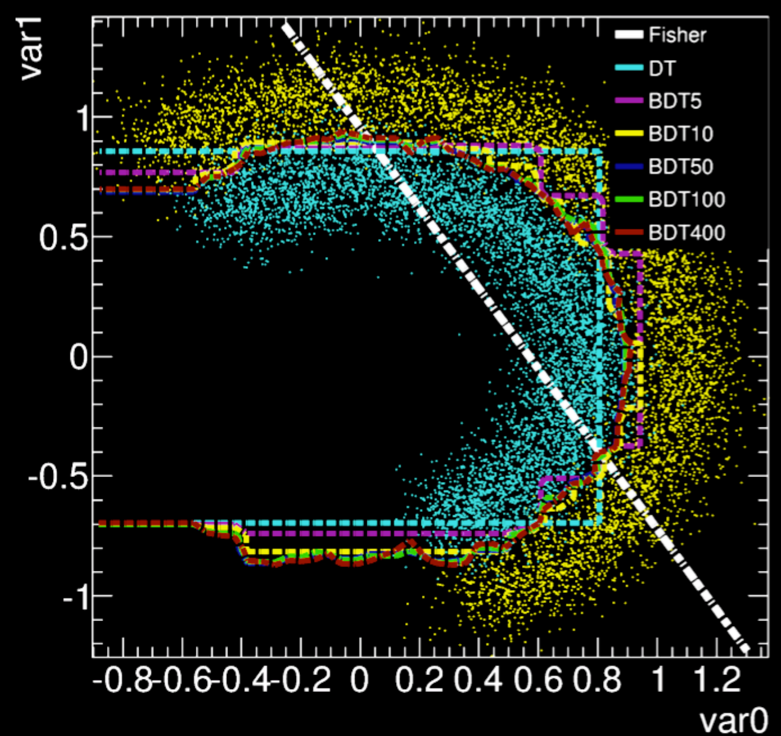




Y. Coadou. European School of Instrumentation in Particle and Astroparticle Physics, Feb. 2016.

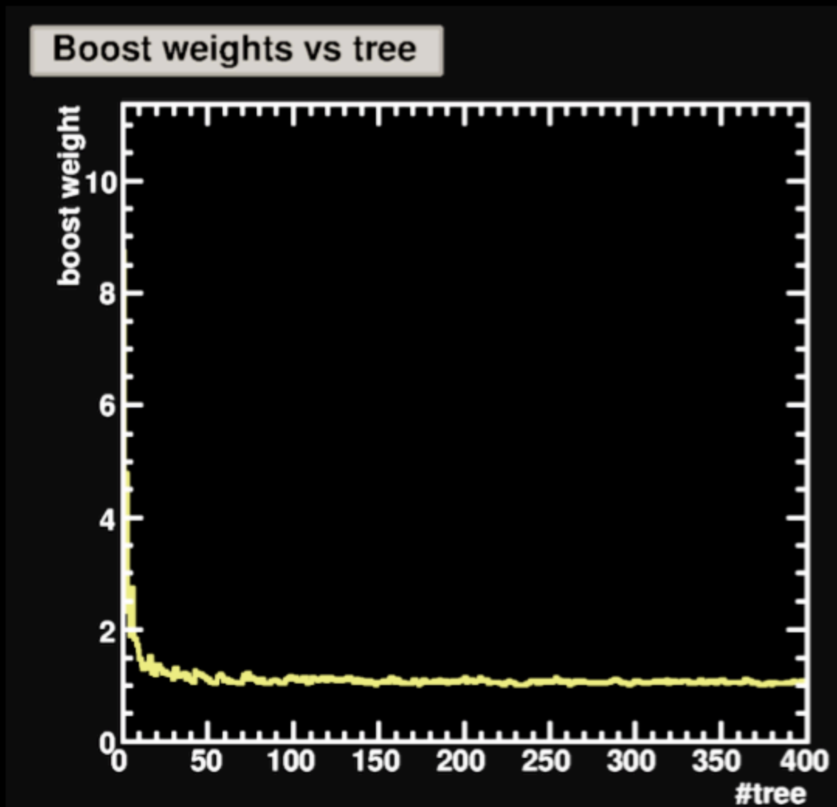
Boosting Walkthrough

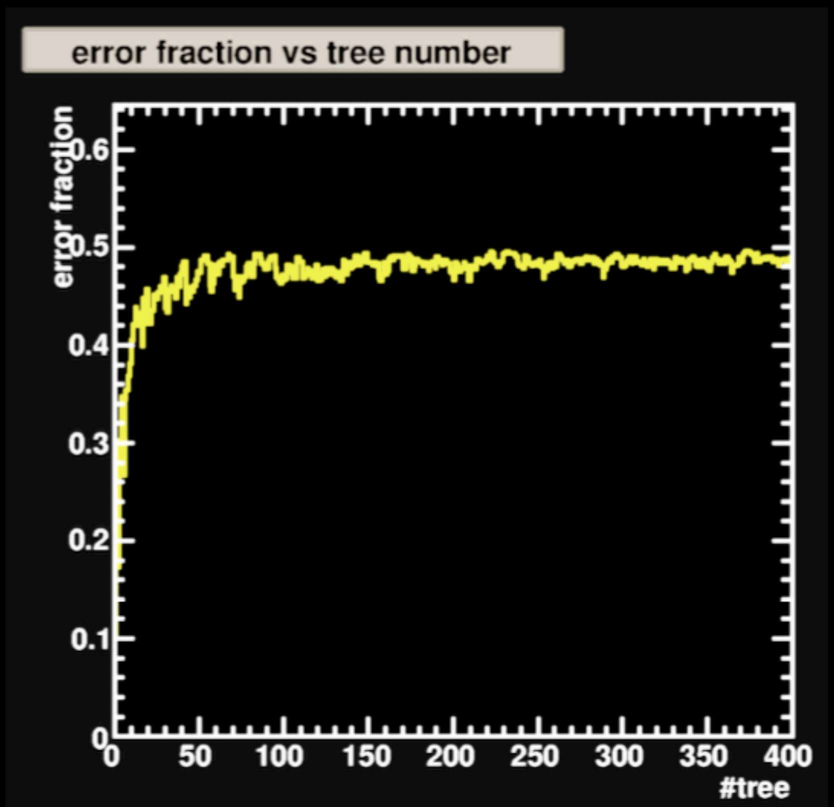




Y. Coadou. European School of Instrumentation in Particle and Astroparticle Physics, Feb. 2016.

Boosting Walkthrough





Y. Coadou. European School of Instrumentation in Particle and Astroparticle Physics, Feb. 2016.

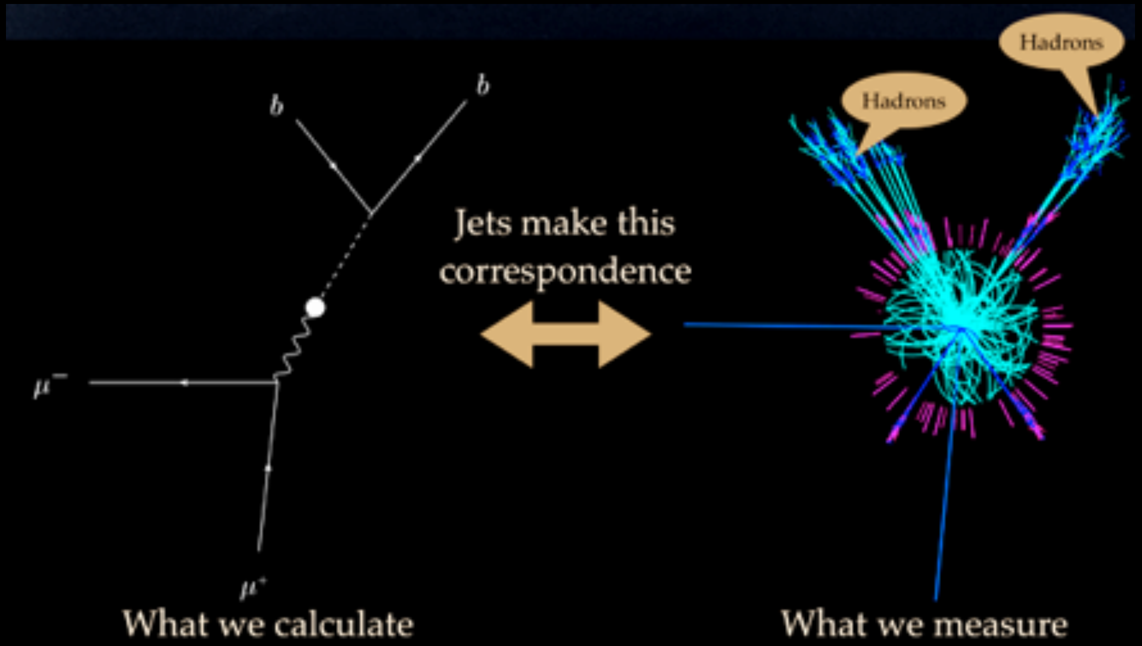
BDTs in Particle Physics

- ATLAS
- CMS
- MiniBooNE
- D0
- BaBar
- LHCb

BDTs in Particle Physics

- ATLAS b-tagging
- CMS
- MiniBooNE
- D0
- BaBar
- LHCb

Need For *b*-tagging



F. Tanedo. http://www.quantumdiaries.org/2011/04/22/when-youre-a-jet-youre-a-jet-all-the-way/

- t decay (t -> Wb)
- Higgs decay (h -> $b\bar{b}$)
- CP violation studies
- Beyond the standard model physics

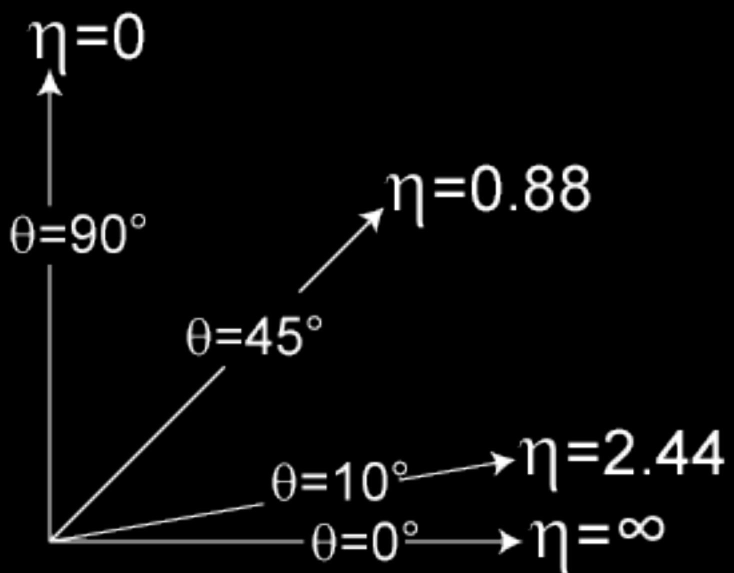
ATLAS Kinematics

$$\eta = -\ln \tan(\theta/2)$$

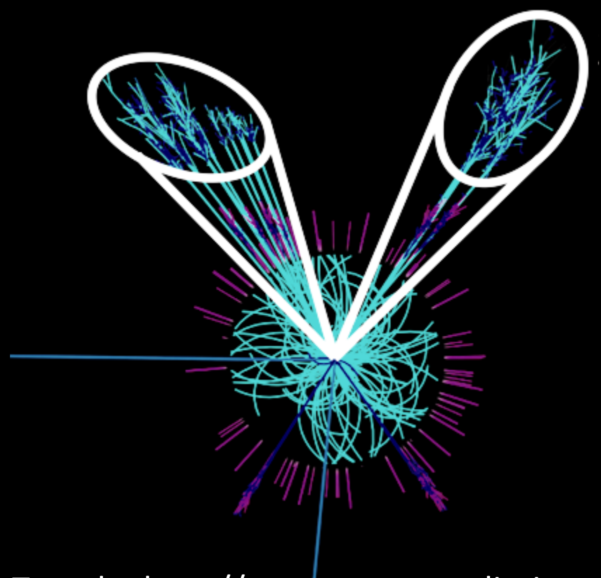
$|\eta|$ < 2.5 detected

$$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

ΔR < 0.3 in jet



https://en.wikipedia.org/wiki/Pseudorapidity

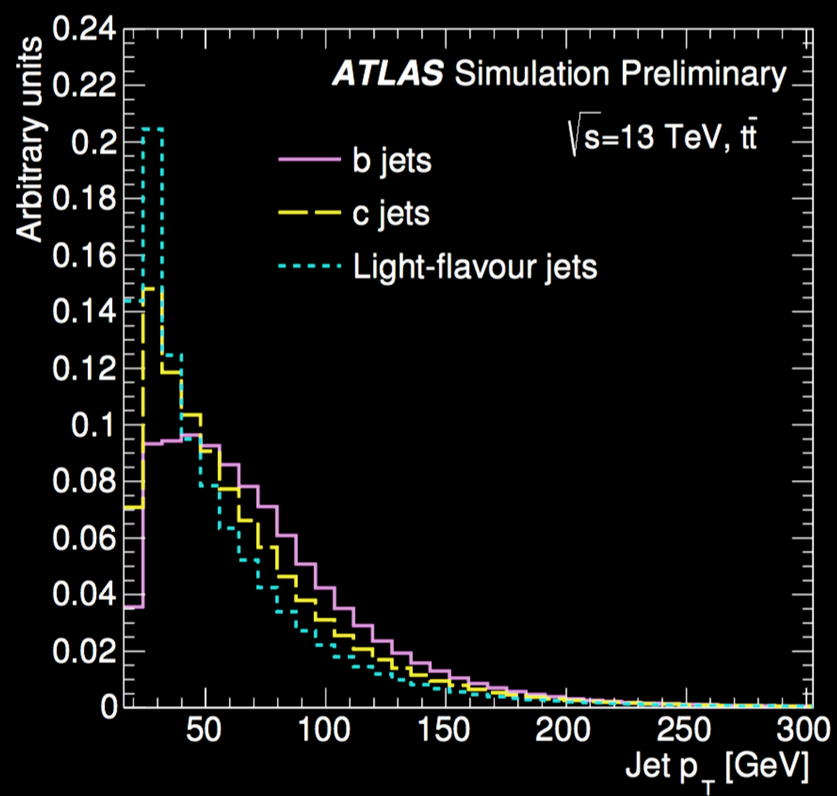


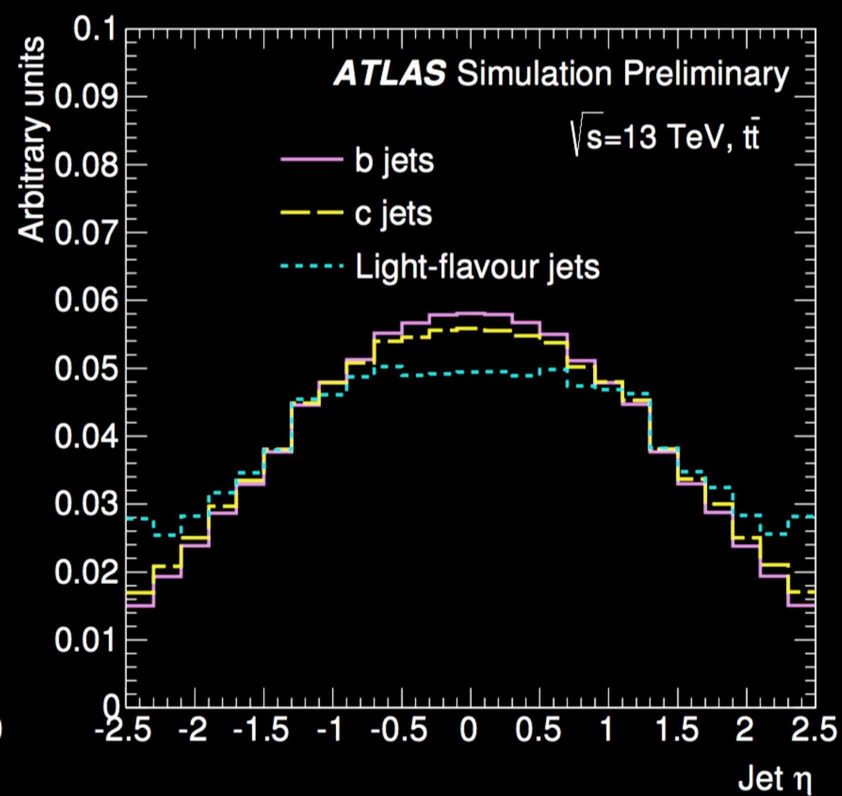
F. Tanedo. http://www.quantumdiaries.org/2011/04/22/when-youre-a-jet-youre-a-jet-all-the-way/

BDT b-tagging in ATLAS

- Model trained on $t\bar{t}$ events corresponding to 13 TeV proton-proton collisions.
  - Trained on *b*-jets as signal and a background of 80% light-flavor jets and 20% *c*-jets.
- Variables are basic kinematic variables η and p_T, and features from 3 reconstruction algorithms.

Basic Jet Observables





ATL-PHYS-PUB-2015-022

BDT Variable List

Input	Variable	Description
Kinematics	$p_T(jet)$	Jet transverse momentum
	$\eta(jet)$	Jet pseudo-rapidity
IP2D, IP3D	$\log(P_b/P_{\mathrm{light}})$	Likelihood ratio between the *b*- and light jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the *b*- and *c*-jet hypotheses
	$\log(P_c/P_{\mathrm{light}})$	Likelihood ratio between the *c*- and light jet hypotheses
SV	$m(\mathrm{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
	$f_{\mathrm{E}}(\mathrm{SV})$	Fraction of the charged jet energy in the secondary vertex
	$N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$	Number of tracks used in the secondary vertex
	$N_{\mathrm{2TrkVtx}}(\mathrm{SV})$	Number of two track vertex candidates
	$L_{xy}(\mathrm{SV})$	Transverse distance between the primary and secondary vertices
	$L_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices
	$S_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\mathrm{jet},\mathrm{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
Jet Fitter	$N_{\mathrm{2TrkVtx}}(\mathrm{JF})$	Number of 2-track vertex candidates (prior to decay chain fit)
	$m(\mathrm{JF})$	Invariant mass of tracks from displaced vertices assuming pion masses
	$S_{\mathrm{xyz}}(\mathrm{JF})$	Significance of the average distance between the primary and displaced vertices
	$f_{\mathrm{E}}(\mathrm{JF})$	Fraction of the charged jet energy in the secondary vertices
	$N_{\mathrm{1\text{-}trk\ vertices}}(\mathrm{JF})$	Number of displaced vertices with one track
	$N_{\geq 2\text{-}\mathrm{trk\ vertices}}(\mathrm{JF})$	Number of displaced vertices with more than one track
	$N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{\mathrm{jet}},\vec{p}_{\mathrm{vtx}})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

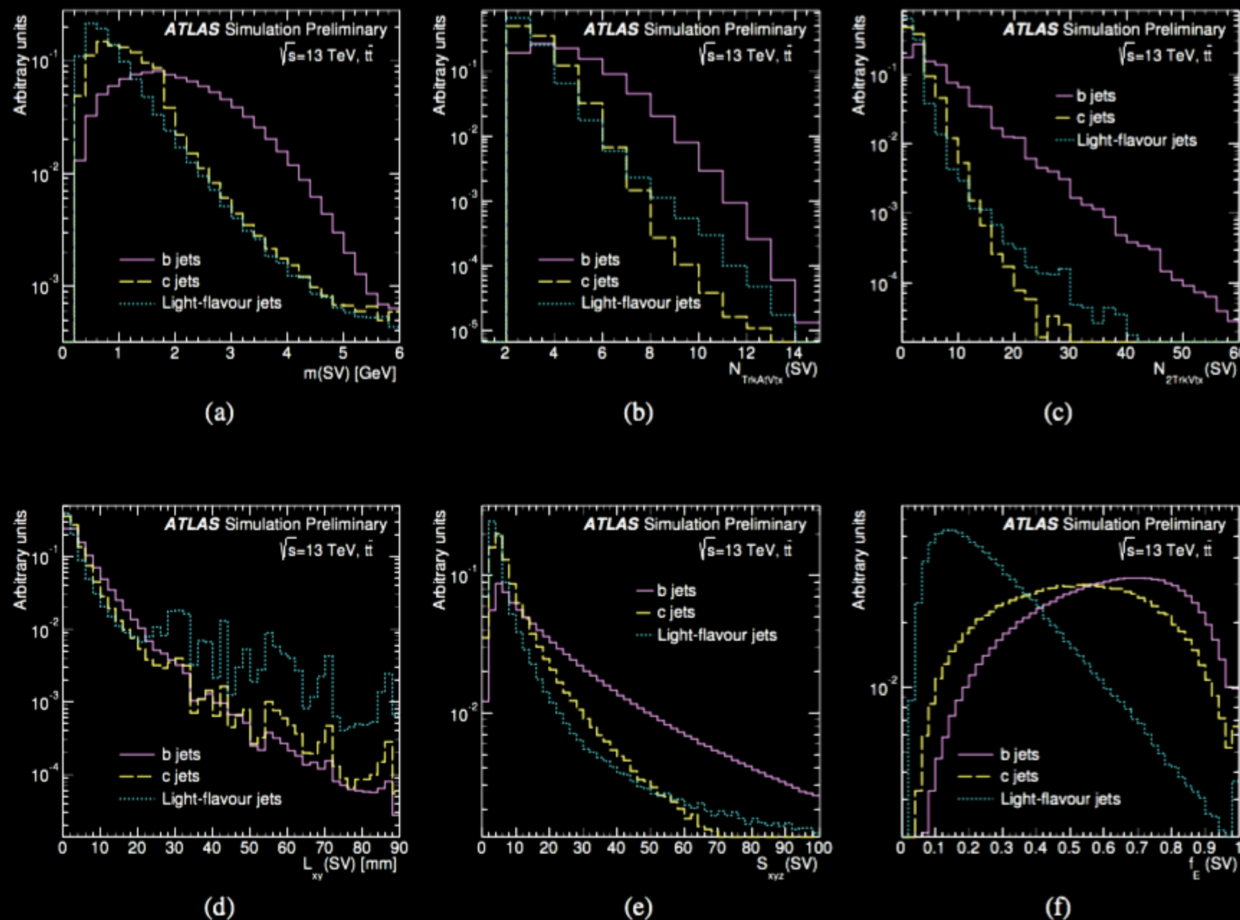
Impact parameter based algorithms

Secondary vertex finding algorithm

Decay chain multi-vertex algorithm

ATL-PHYS-PUB-2015-022

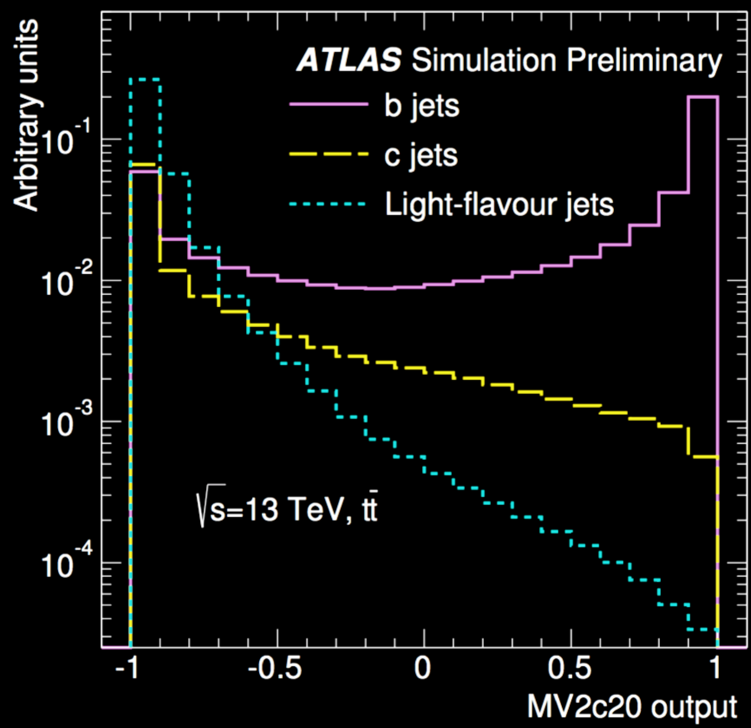
SV Algorithm Variables



ATL-PHYS-PUB-2015-022

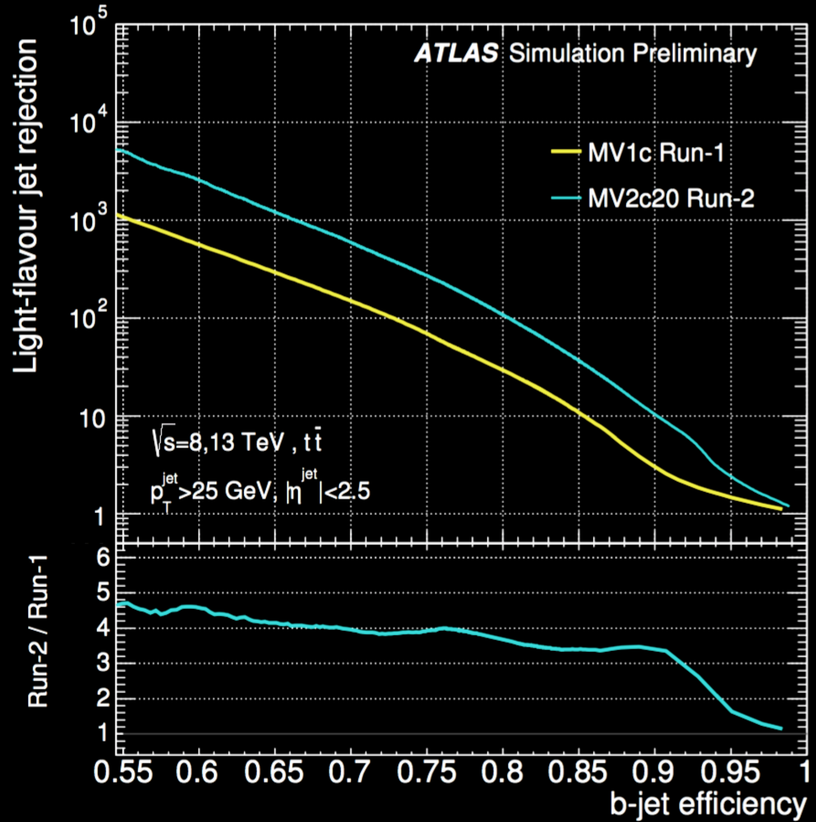
Predicted Performance

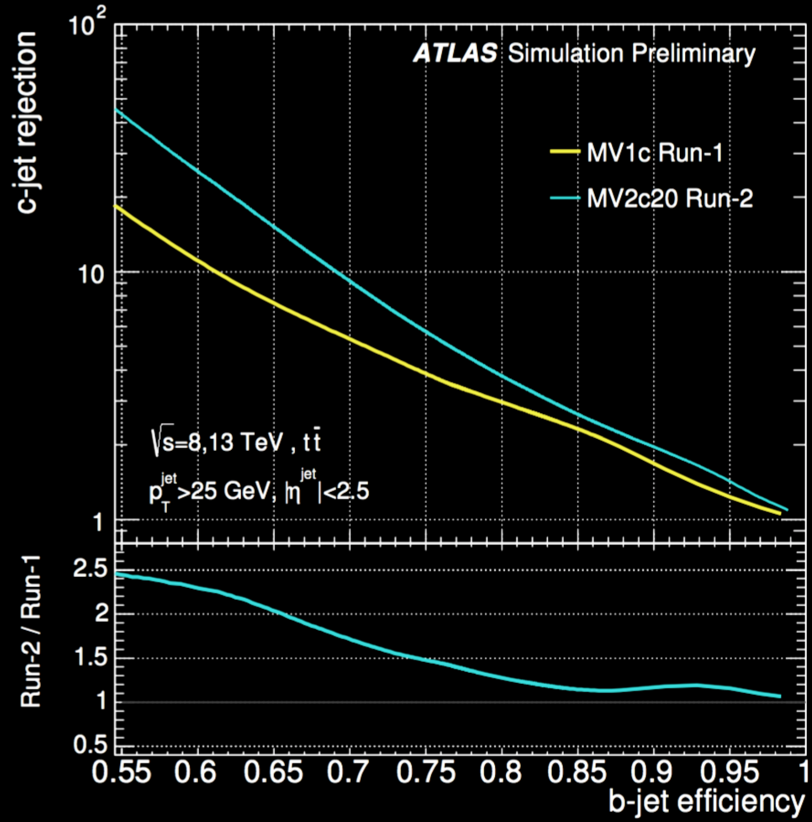
ATL-PHYS-PUB-2015-022



Cut Value	b-jet Efficiency [%]	c-jet Rejection	τ-jet Rejection	Light-jet Rejection
0.4496	60	21	93	1900
-0.0436	70	8.1	26	440
-0.4434	77	4.5	10	140
-0.7887	85	2.6	3.8	28

Predicted Improvement





- Some hardware, some software

ATL-PHYS-PUB-2015-022

Final Thoughts

- Decision trees are fairly weak classifiers that are easy to understand.
- Boosting turns them into stronger classifiers but makes them more mysterious.
- BDTs show up a lot in particle physics.
- Propagating uncertainties difficult.
- <u>Still need to convince people your model produces good results.</u>

References

- T. G. Dietterich. Machine learning research: Four current directions. AI Magazine, 18(4):97–136, 1997.
- G. Cowan. Multivariate statistical methods and data mining in particle physics. CERN Academic Training Lectures, June 2008.
- ATLAS Collaboration. Expected performance of the ATLAS b-tagging algorithms in run-2. ATLAS Note, (ATL-PHYS-PUB-2015-022), July 2015.
- Y. Coadou. Boosted decision trees. European School of Instrumentation in Particle and Astroparticle Physics, February 2016.
- G. James, D. Witten, T. Hastie, and R. Tibshirani. An Introduction to Statistical Learning. Springer Texts in Statistics. Springer, 7th printing edition, 2017.
- D. Koshelenko. Boosting b-tagging performance at high transverse momentum. CERN Summer Student Session, July 2017.