Application of gradient boosting in the kinematic reconstruction of ttbar events

Kolloquium in the program Computing in Science

Karim Ritter von Merkl

28. September 2020

Standard model and beyond?

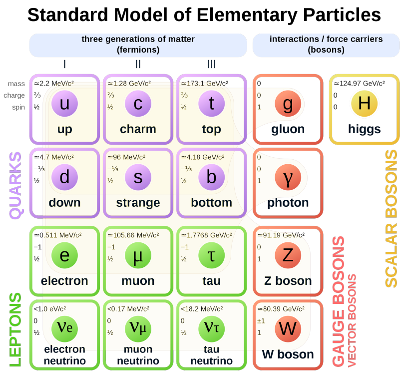


Figure: Fundamental particles of the standard model

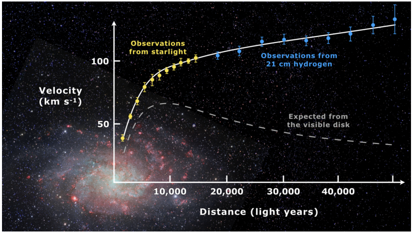


Figure: Predicted and observed rotational velocities of a spiral galaxy

Top Quarks

- heaviest known fundamental particle
  - 173 GeV $\approx m_{\text{gold}}$
- Coupling to Higgs boson close to one
  - chance or BSM physics?
- decays before hadronization

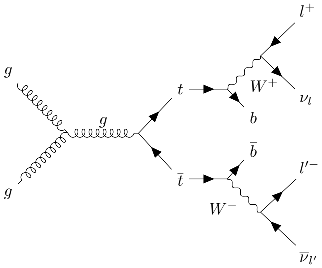


Figure: Production and dileptonical decay of a $t\bar{t}$ pair

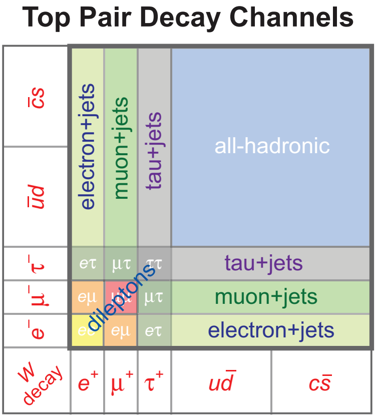


Figure: Diagram showing the relative frequencies of the top-antitop pair decay channels

Reconstruction via an analytical solution (Sonnenschein)

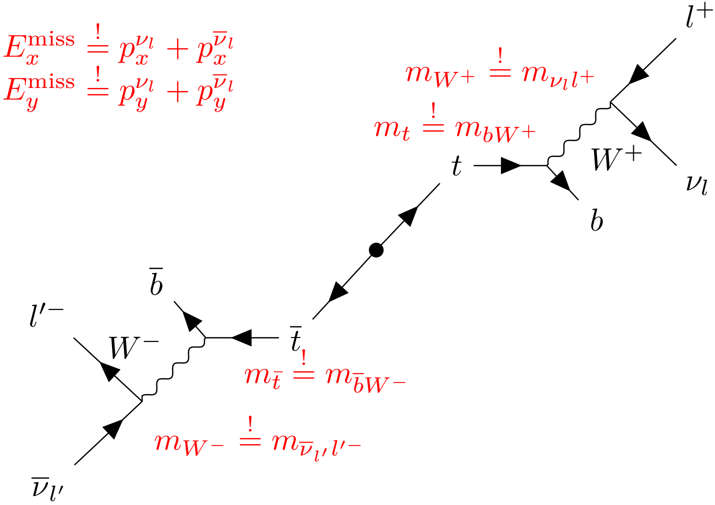


Figure: Dileptonic decay with the additional equations introduced to find an analytical solution

Predictions of regression decision trees

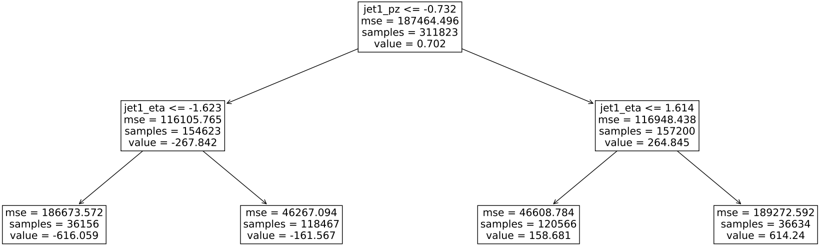


Figure: Example of a decision tree predicting p_z with depth 2

- Every event starts in the root
- If jet1_pz ≤ -0.732, it progresses to the left, else to the right
- Path ends in a leaf. Value of that leaf is predicted
- Building algorithm determines splits and values

Decision tree building algorithm

- value: mean p_z of all events in that node
- mse: variance
- choose split with highest variance decrease ΔI
- feature importance of a variable: fraction of total variance decrease by splits on that feature
- For gradient boosting: maximize $\Delta\tilde{I}$

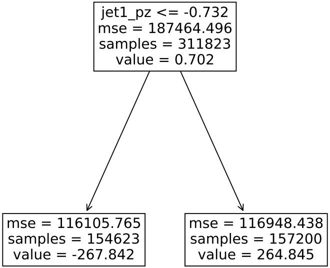


Figure: Decision tree predicting p_z of depth 1

$$\Delta I = \frac{N_{\text{node}}}{N_{\text{total}}}\left(I_{\text{node}} - \frac{N_{\text{right}}}{N_{\text{node}}} I_{\text{right}} - \frac{N_{\text{left}}}{N_{\text{node}}} I_{\text{left}}\right)$$

$$\Delta\tilde{I} = \frac{N_{\text{left}} N_{\text{right}}}{N_{\text{left}} + N_{\text{right}}}(\bar{y}_{\text{left}} - \bar{y}_{\text{right}})^2$$

The gradient boosting algorithm

- input x_i, target y_i, prediction $\hat{y}_i$
- final prediction: sum of individual predictions
- minimize loss function by training next tree
- Here: Using least square loss $L(\hat{y}_i, y_i) = (y_i - \hat{y}_i)^2$
- train next tree trained on the error made so far

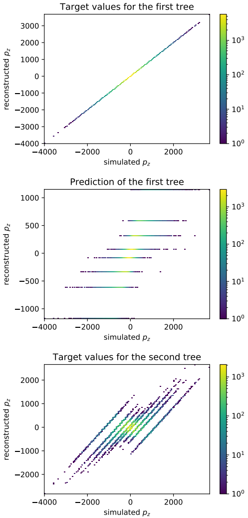


Figure: First step of training GBDTs to predict p_z

Data sets

- Approximate detector simulation
  - selfgenerated using Delphes
- Full detector simulation
  - centrally produced NanoAOD
- 2 btags, 2 leptons opposite charge and $p_T > 10\,\text{GeV}$
- 4 sets: Delphes (sorted/unsorted), NanoAOD (sorted/unsorted)

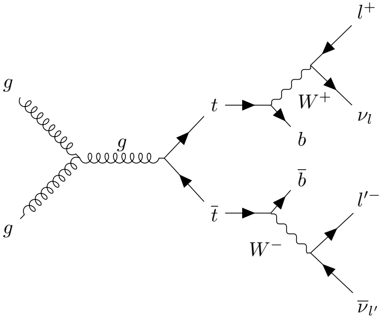


Figure: top pair production via gluon fusion and dileptonic decay

- lept1 positive, lept2 negative
- unsorted sets: jet1 higher p_T
- Delphes sorted: jet1 minimizes $|p_x^{\text{top}} - p_x^{\text{jet1}} - p_x^{\text{lept1}} - p_x^{\text{neutrino}}|$
- NanoAOD sorted: use partonFlavour to find b/antib jets

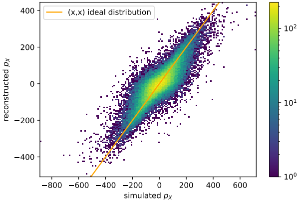
Input to training

Type	shortcut	variables
lepton	lept	pt, eta, phi, px, py, pz, E
jet	jet	pt, eta, phi, mass, px, py, pz, E, btagDeepB (NanoAOD)
MET	MET	E_T^{miss}, phi, px, py

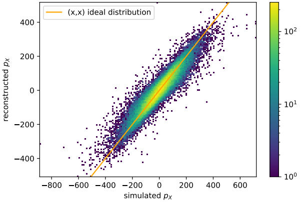
Table: Overview of the input variables used for the reconstruction.

- input: kinematic variables of lepton, jets and MET
- includes redundancies, might be more accessible for a decision trees
- train models with different parameters
- choose the one with the highest $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

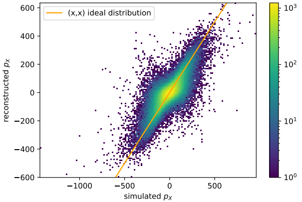
Resulting reconstruction



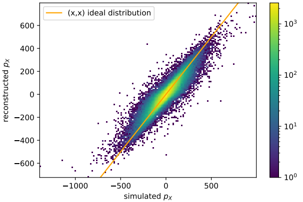
(a) Delphes unsorted



(b) Delphes sorted



(c) NanoAOD unsorted



(d) NanoAOD sorted

Figure: Two dimensional histogram of points $(p_{x,\text{top}}, p^{\text{predicted}}_{x,\text{top}})$ for each data set.

Differences between variables and sets

- R^2 on test set: $p_z > p_x \approx p_y > p_T$
- performance on p_x, p_y gain a lot from sorting
- performance on p_z, p_T gain less from sorting
- on sorted sets about equally good
- Generally: Delphes unsorted > NanoAOD unsorted

Feature importance

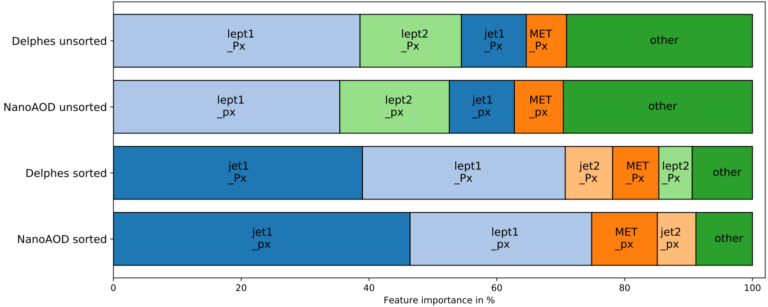


Figure: Feature importance best model per data set predicting p_x

		jet1	jet2	lept1	lept2
Delphes	unsorted	0.0634	0.0197	0.6476	-0.5040
	sorted	0.6826	-0.5692	0.6476	-0.5040
NanoAOD	unsorted	0.0851	0.0169	0.5807	-0.4596
	sorted	0.6659	-0.5265	0.5761	-0.4545

Table: Correlation of columns p_x with top's p_x

Summary

Conclusion:

- powerful and comprehensible reconstruction method
- search for improvements possible by analyzing learning
- Separating bjets and antibjets helpful but not crucial
- Correlating variables turned out to be very helpful

Outlook:

- feature selection and dimensionality reduction
  - remove redundant features, inroduce new ones
- evaluate effect of cuts
  - kinematic cuts seem helpful so far
- more efficient implementation
  - robust model using more data
- compare with analytical solution
- test application in BSM theories

Backup

Proton collisions at the Large Hadron Collider

- proton-proton collider at 13 TeV center of mass energy
- protons not fundamental
- mixture of quarks and gluons (partons)
- proton momentum split among partons
- proton collision = collision of different partons
- rest frame of proton collision $\neq$ rest frame of parton collision

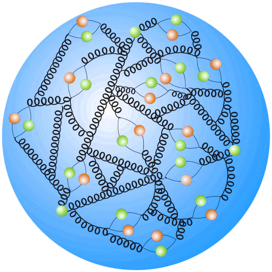


Figure: Schematic representation of the structure of the proton

Coordinates

- polar coordinates in the transverse plane are natural due to symmetry
- polar angle θ not invariant under boosts
- pseudo rapidity $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$ transforms additively for highly relativistic particles
  - differences invariant under boosts

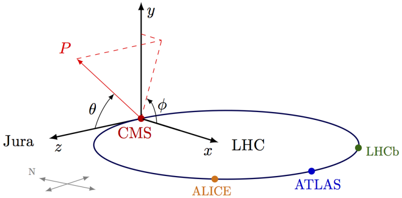


Figure: Coordinate system within a detector