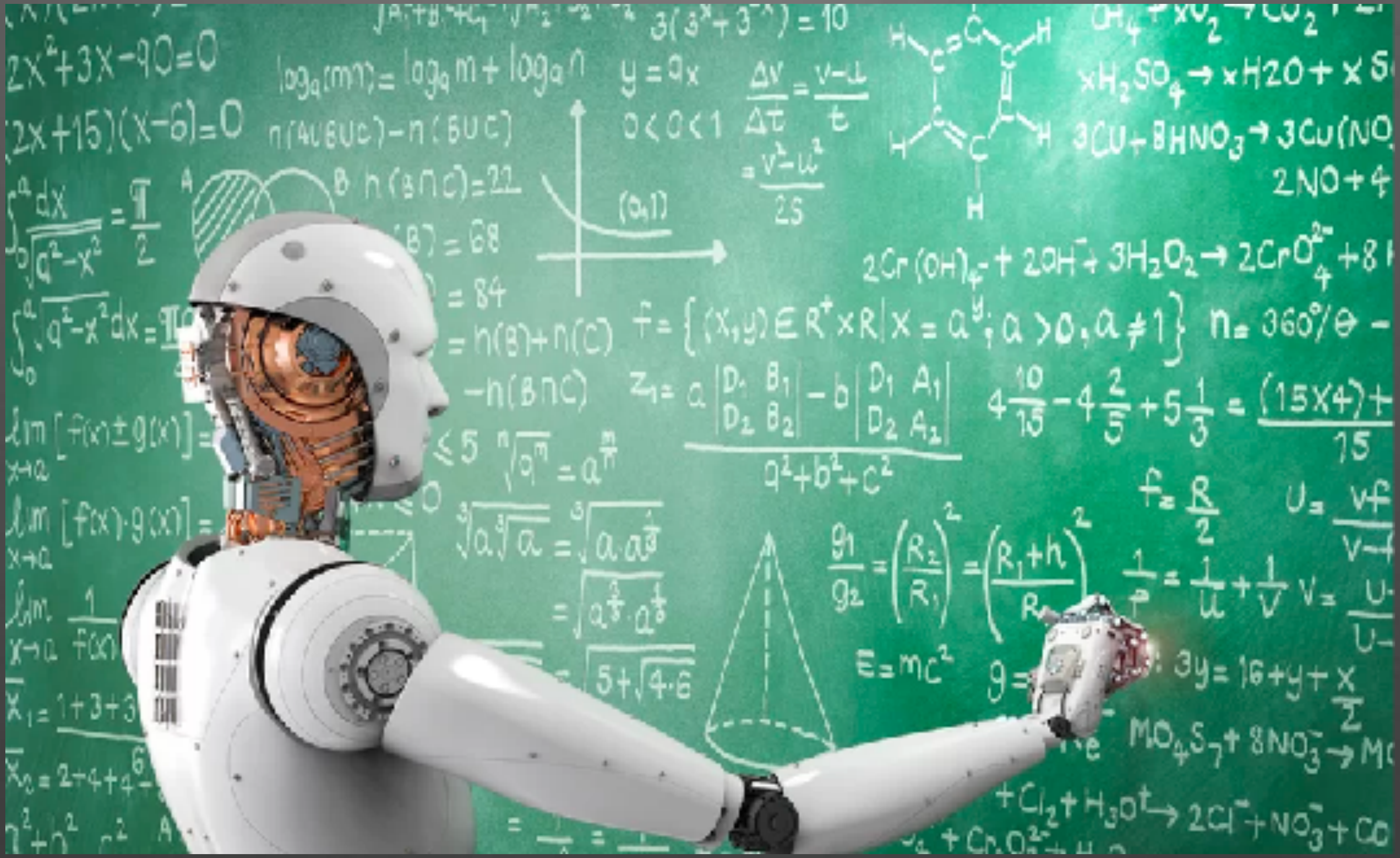
Learning Physics from Machine Learning



Credit: Phonlamai Photos

Spencer Chang (U. Oregon, NTU)

HKUST IAS Theory - Physics Opportunities and Advanced Tools 1/11/19

in collaboration with T. Cohen, B. Ostdiek

Outline

Outline

- Machine learning: Effective, but not understandable

Outline

- Machine learning: Effective, but not understandable
- Data planing to diagnose importance of physics variables

Outline

- Machine learning: Effective, but not understandable
- Data planing to diagnose importance of physics variables
- A simple collider application

Outline

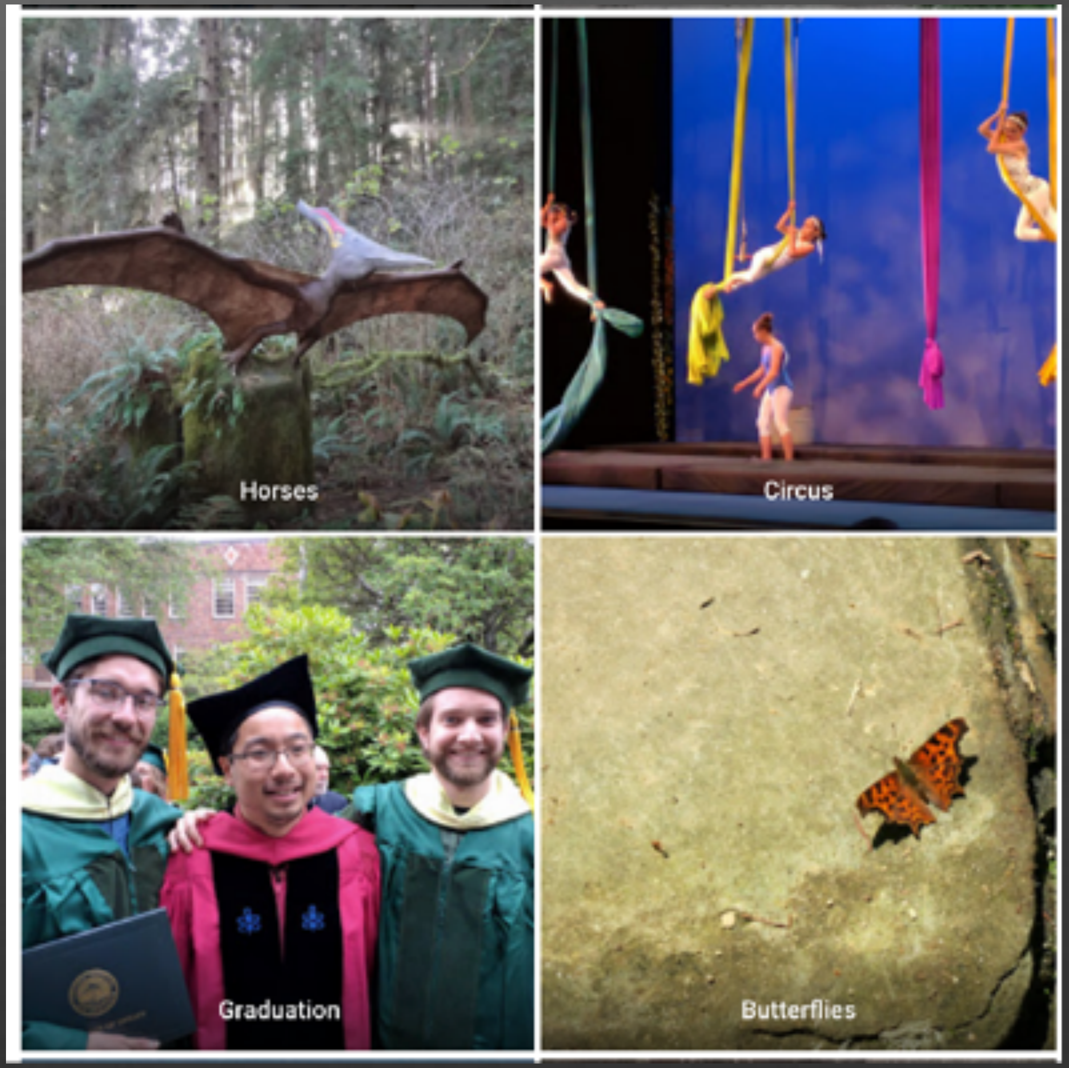
- Machine learning: Effective, but not understandable
- Data planing to diagnose importance of physics variables
- A simple collider application
- Future Directions

Machine Learning Breakthroughs

DeepMind

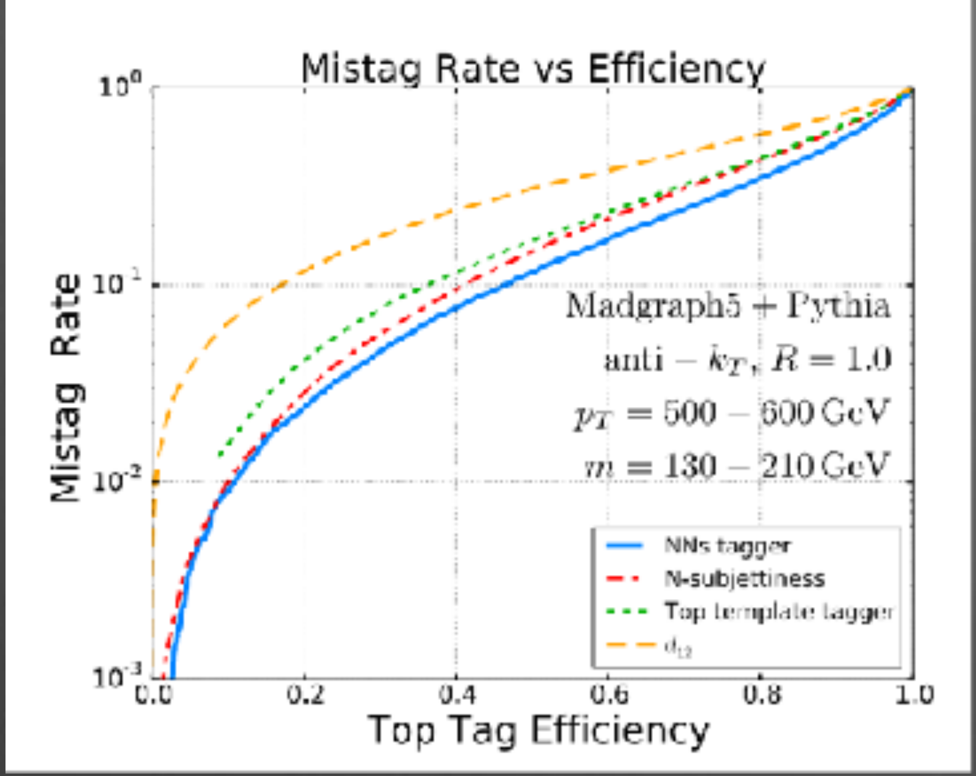
Generating fake photos (NVIDIA)

Image classification (Google Photos)

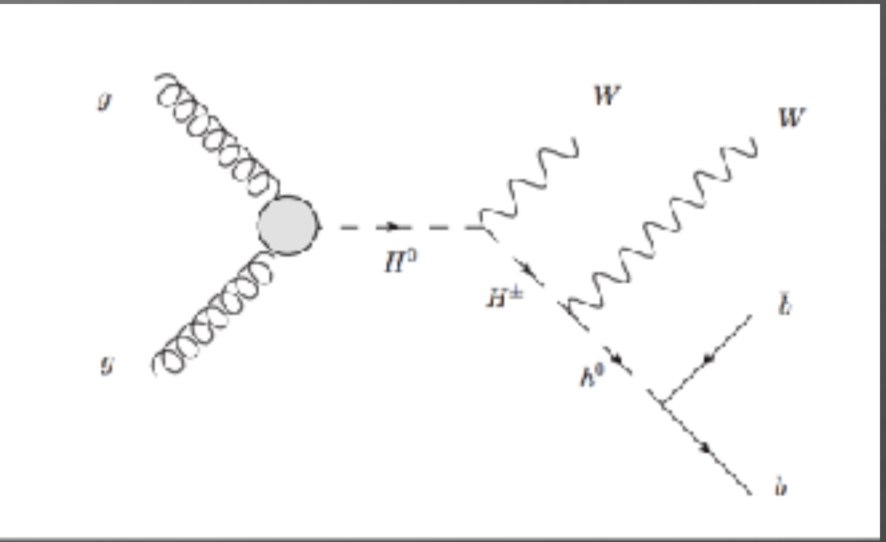


Understanding Machine Learning

Recent Progress in Machine Learning in High Energy Physics



Better Selection Algorithm for energetic top quarks
Almeida et.al. 1501.05968



Deep Learning for Beyond the Standard Model signals
Baldi et.al. 1402.4735

CaloGAN
Fast Detector Simulation
Paganini et.al. 1712.10321

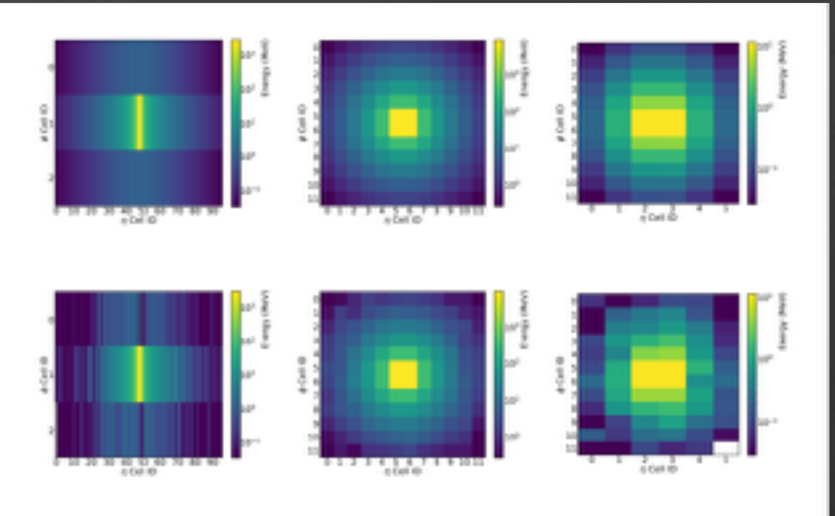
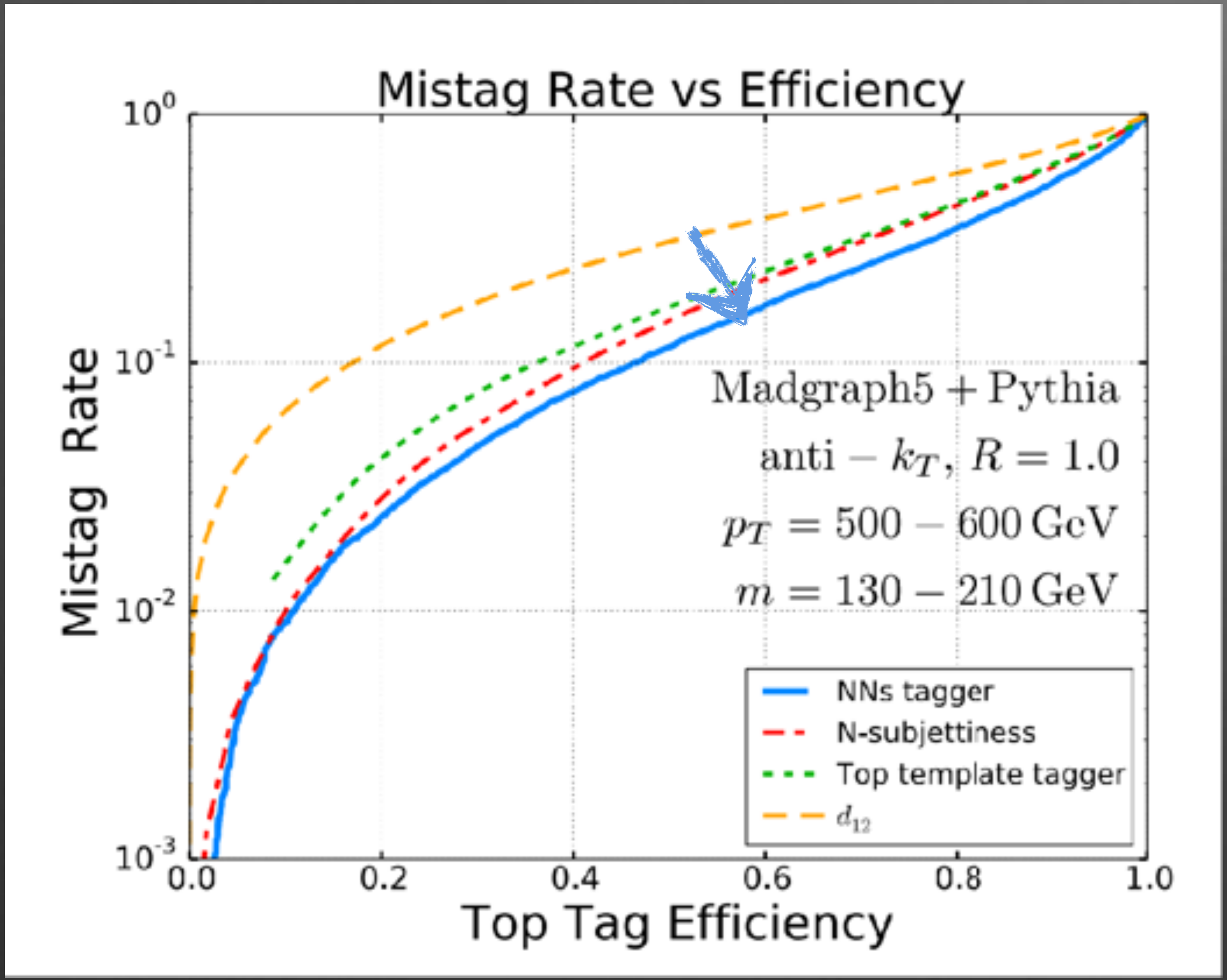


FIG. 6: Average e^+ GEANT4 shower (top), and average e^+ CALOGAN shower (bottom), with progressive calorimeter depth (left to right).

Learning Physics

Focus On Classification

Clear that machine learning is better at classifying than human experts



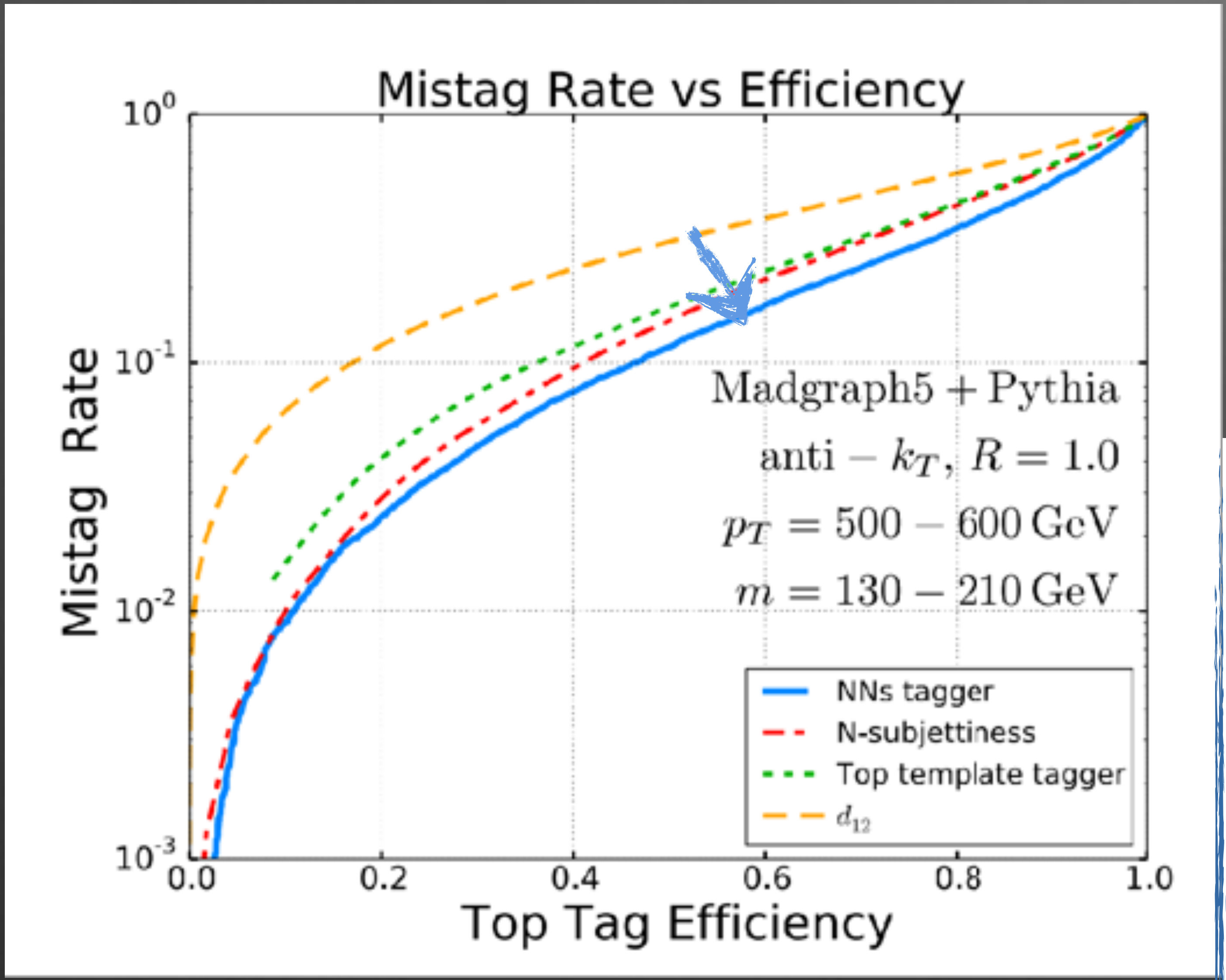
Top quark selection

Almeida et.al. 1501.05968

Learning Physics

Focus On Classification

Clear that machine learning is better at classifying than human experts



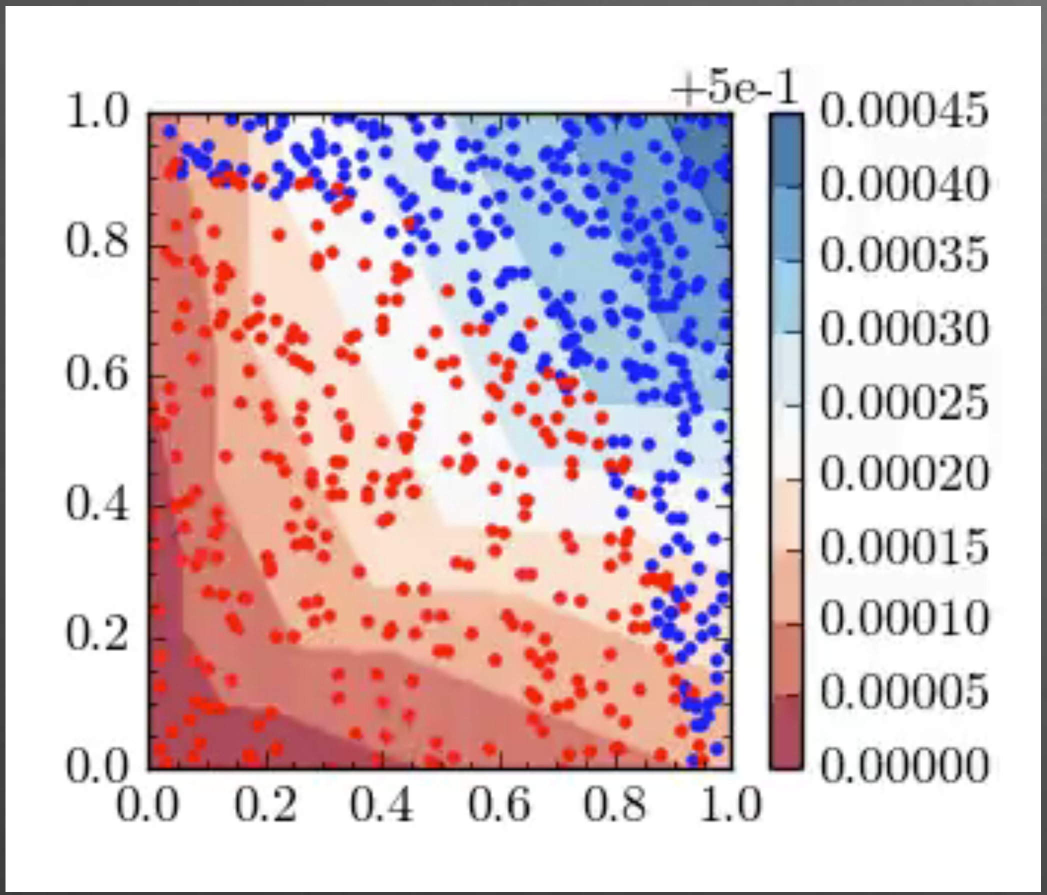
Top quark selection
Almeida et.al. 1501.05968

But can we learn physics they are utilizing? (e.g. what new variables or correlations are useful?) Important for theorists and experimentalists!

Understanding Classifiers

1709.10106 PRD SC w/ Cohen, Ostdiek

y

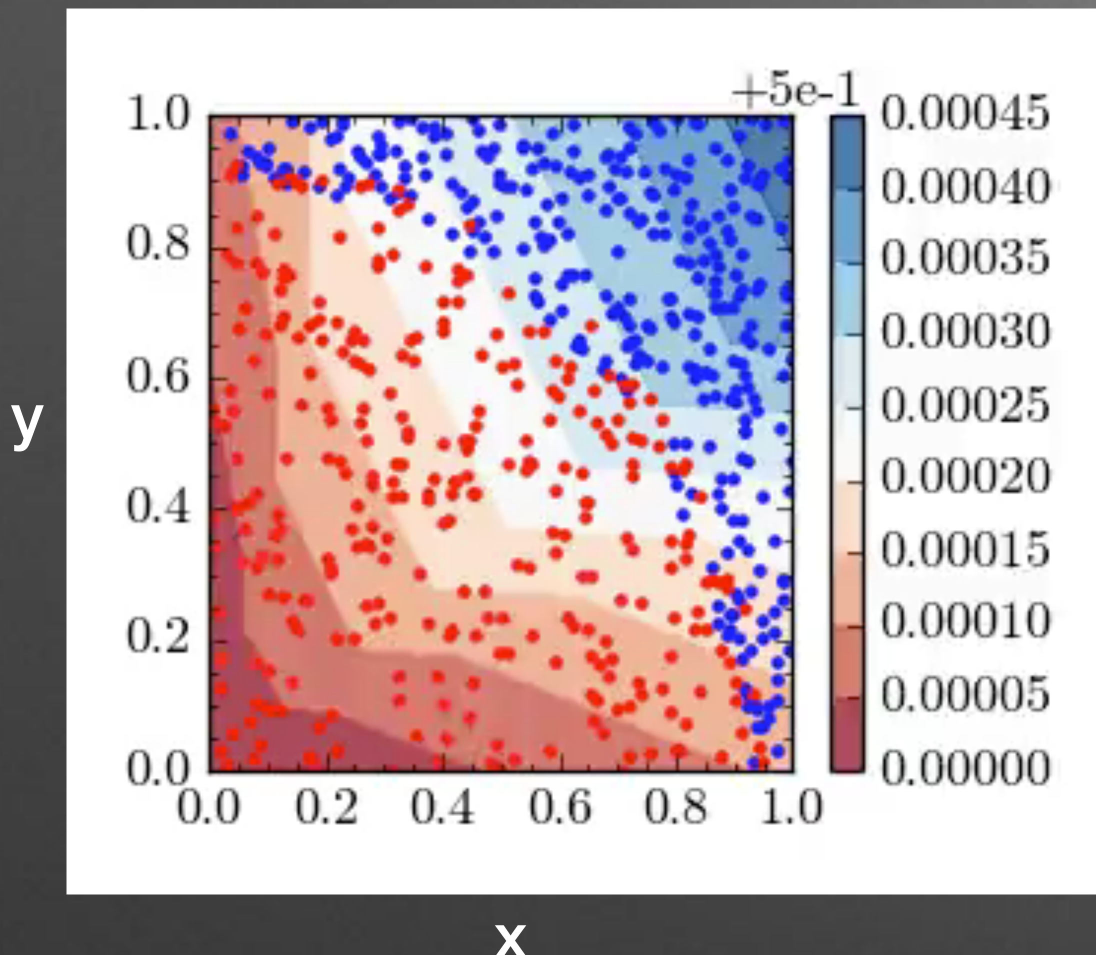


x

Training data

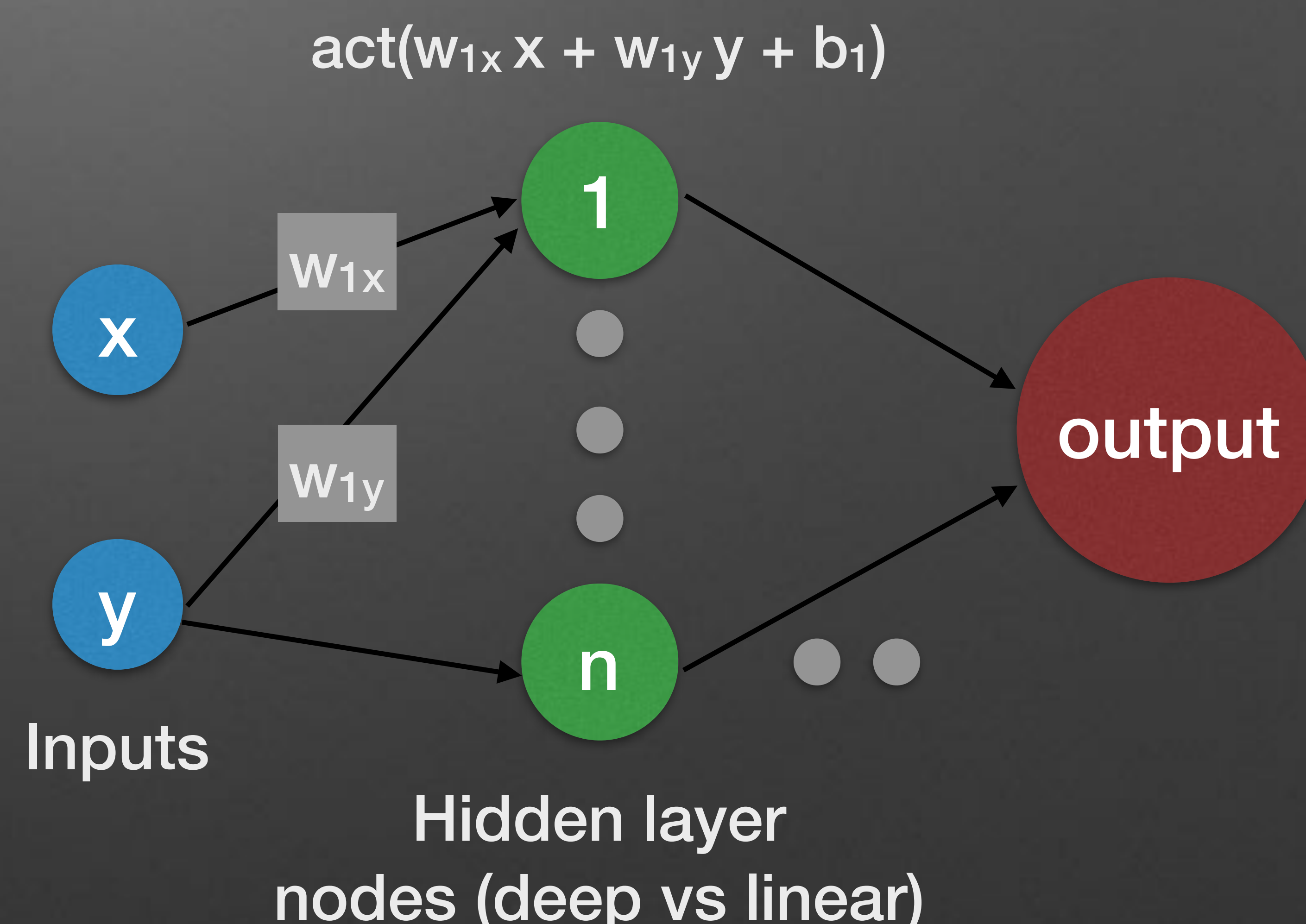
Understanding Classifiers

1709.10106 PRD SC w/ Cohen, Ostdiek



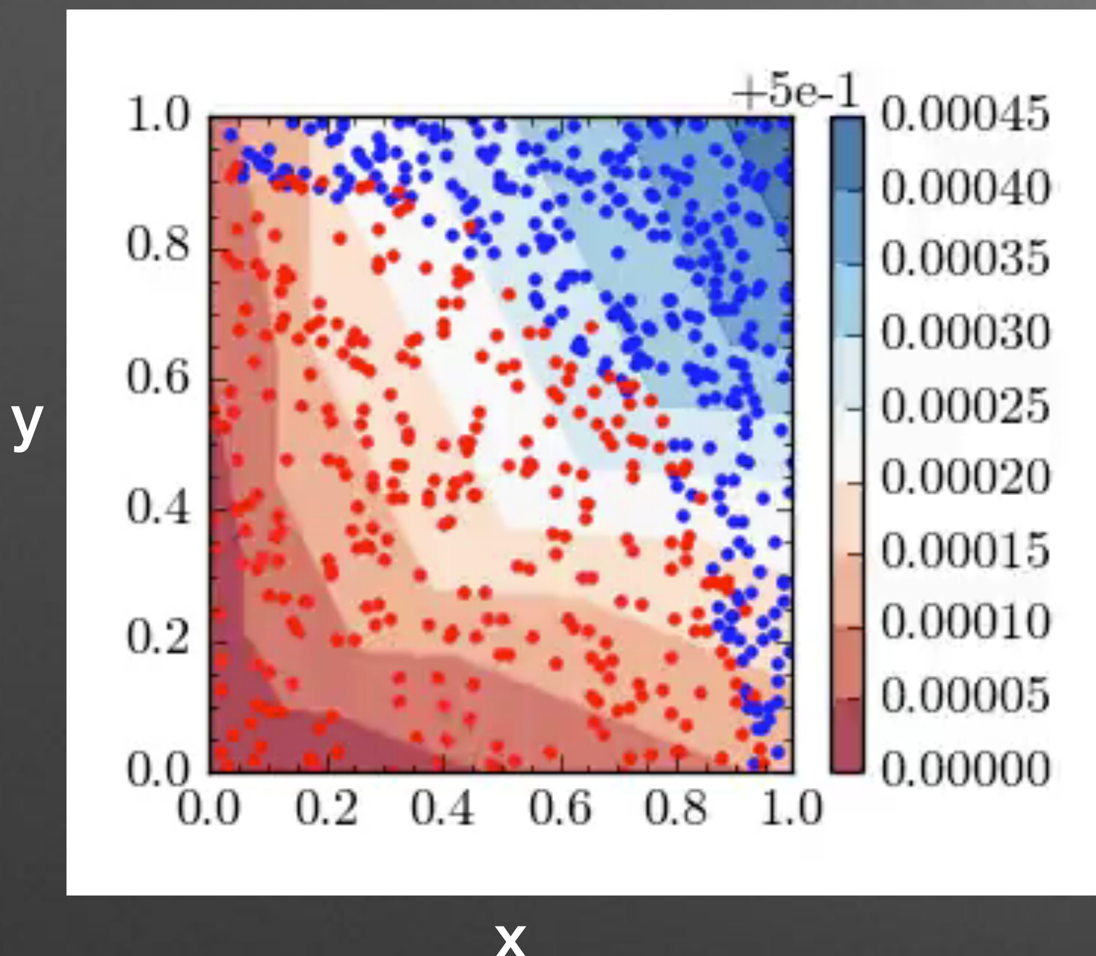
Training data

Neural Networks (NNs) excel at classification (e.g. red vs blue)



Understanding Classifiers

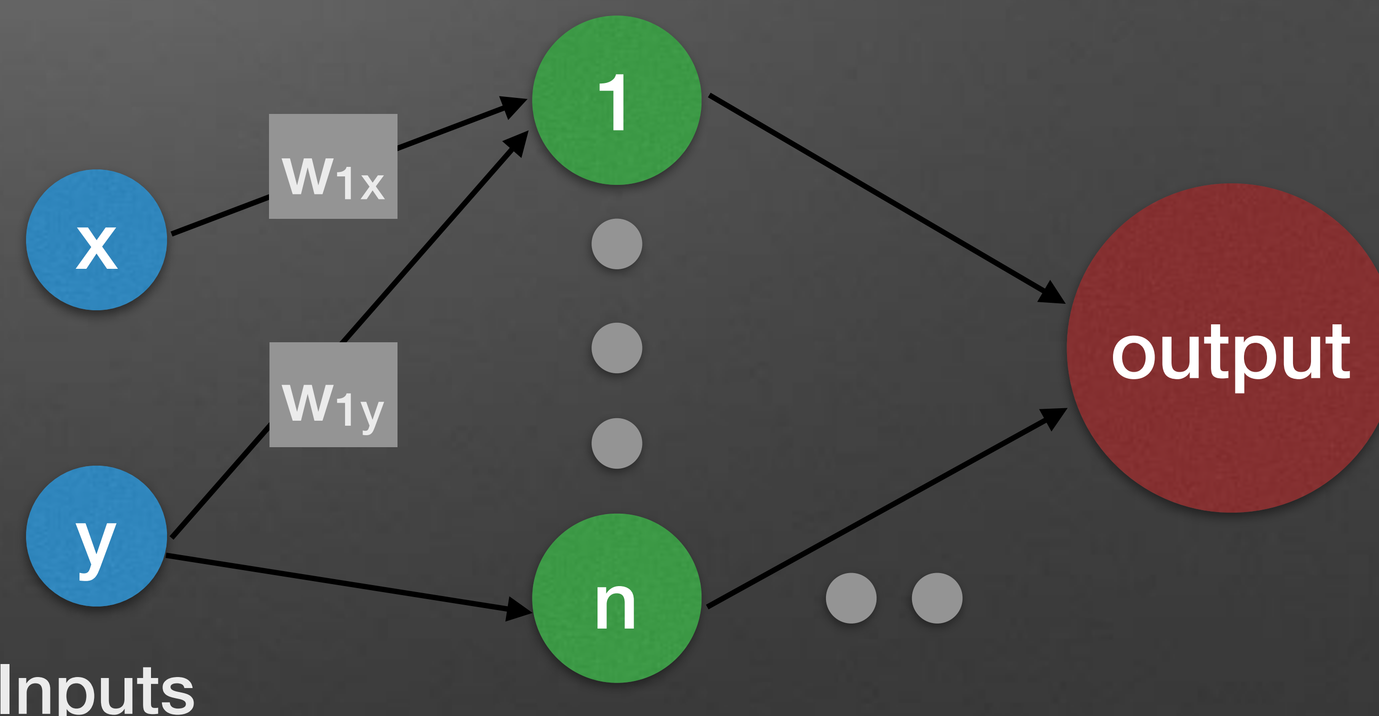
1709.10106 PRD SC w/ Cohen, Ostdiek



Training data

Neural Networks (NNs) excel at classification (e.g. red vs blue)

$\mathrm{act}(w_{1x}\, x + w_{1y}\, y + b_1)$

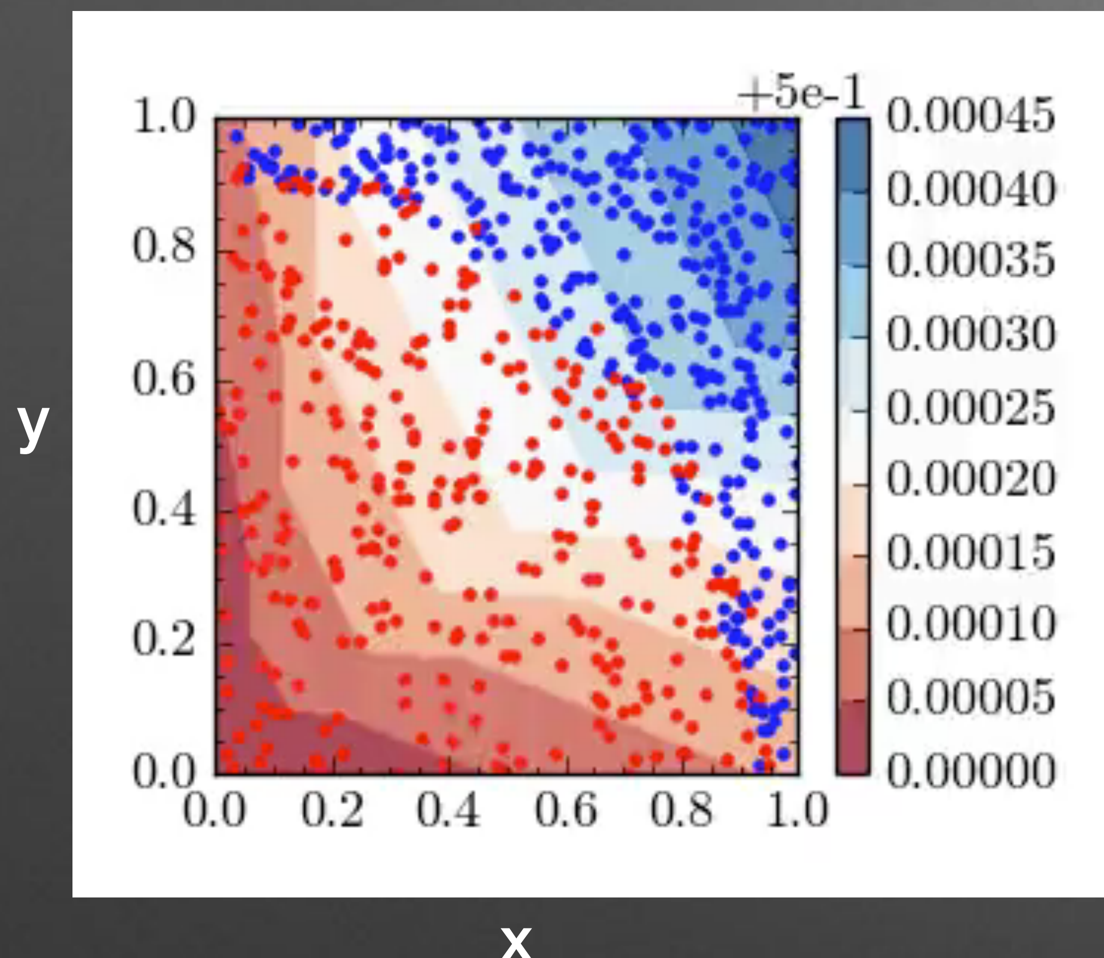


Hidden layer
nodes (deep vs linear)

Deep NN can approximate any function, weights & biases found by training

Understanding Classifiers

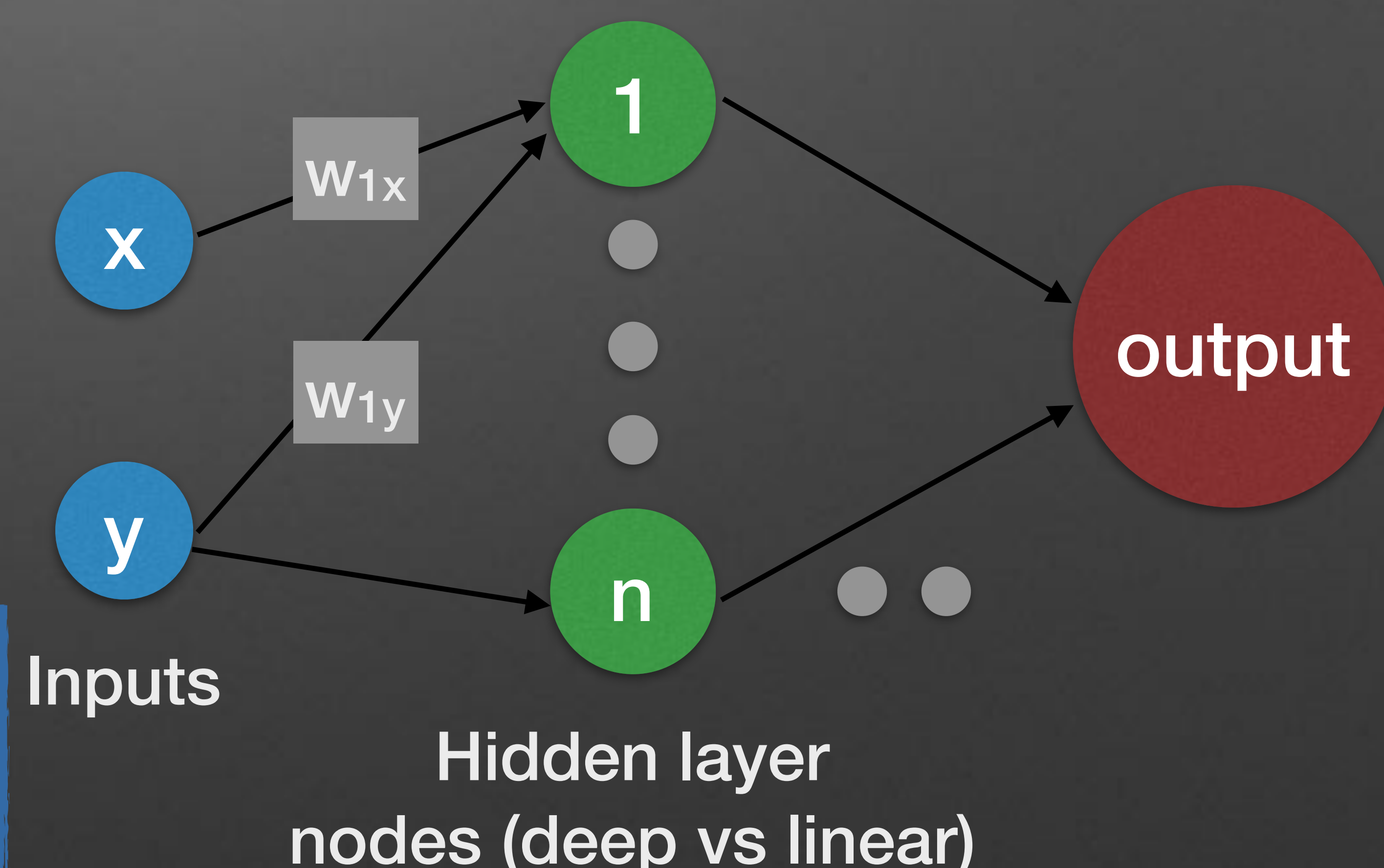
1709.10106 PRD SC w/ Cohen, Ostdiek



Training data

Neural Networks (NNs) excel at classification (e.g. red vs blue)

$act(w_{1x}\,x + w_{1y}\,y + b_1)$

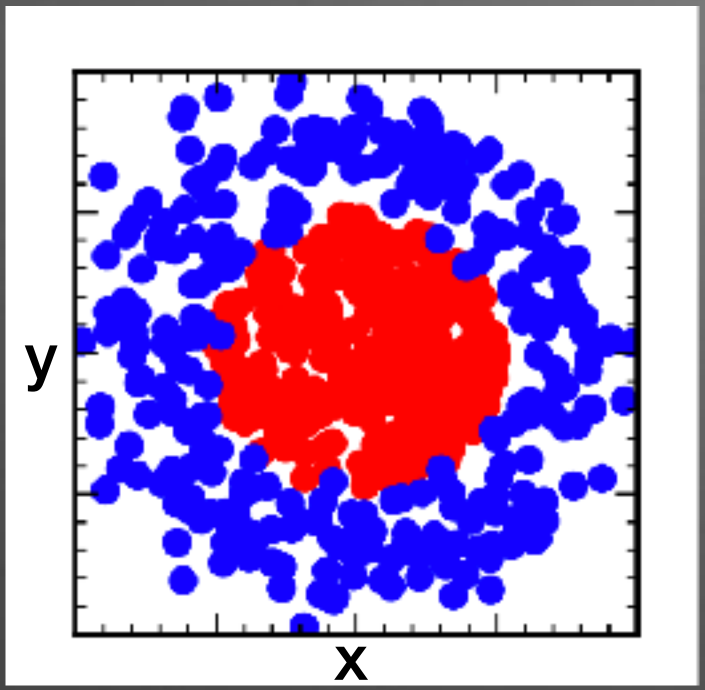


Deep NN can approximate any function, weights & biases found by training

Interpreting NN Classifier

1709.10106 PRD SC w/ Cohen, Ostdiek

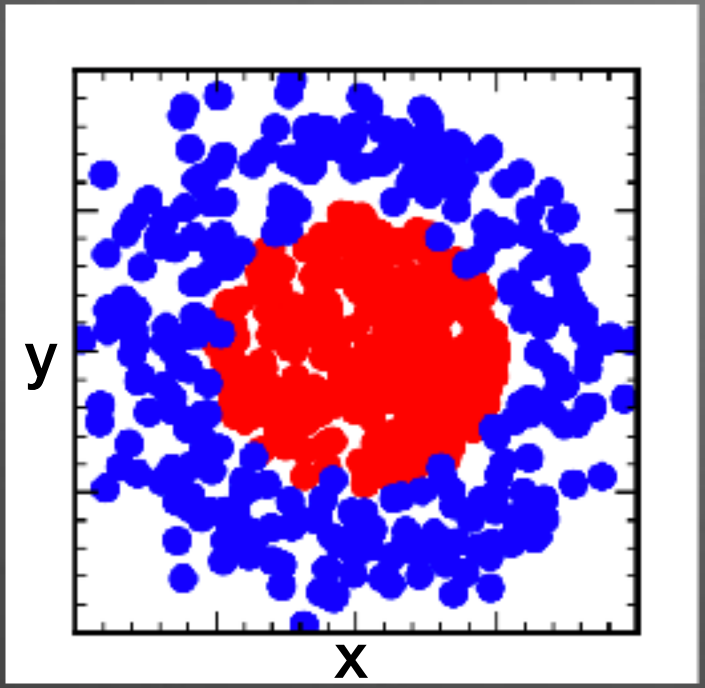
Training data

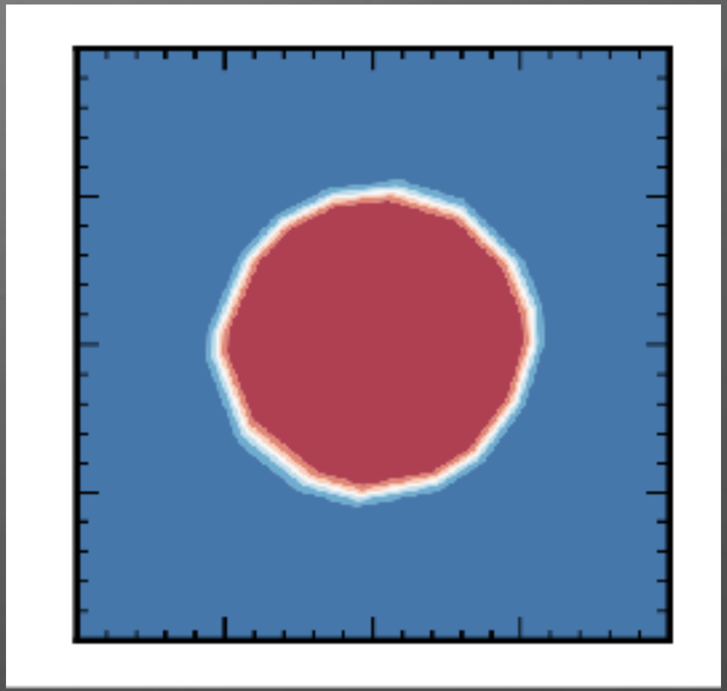


Interpreting NN Classifier

1709.10106 PRD SC w/ Cohen, Ostdiek

Training data



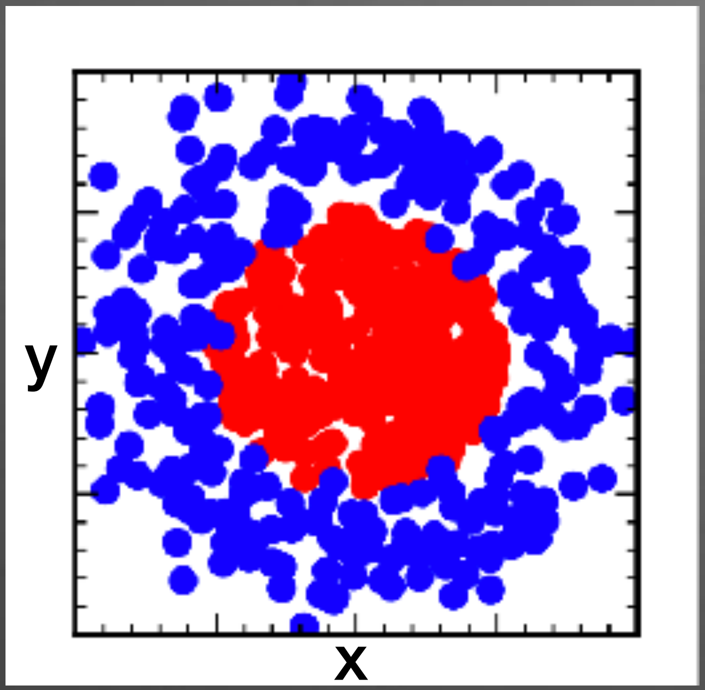


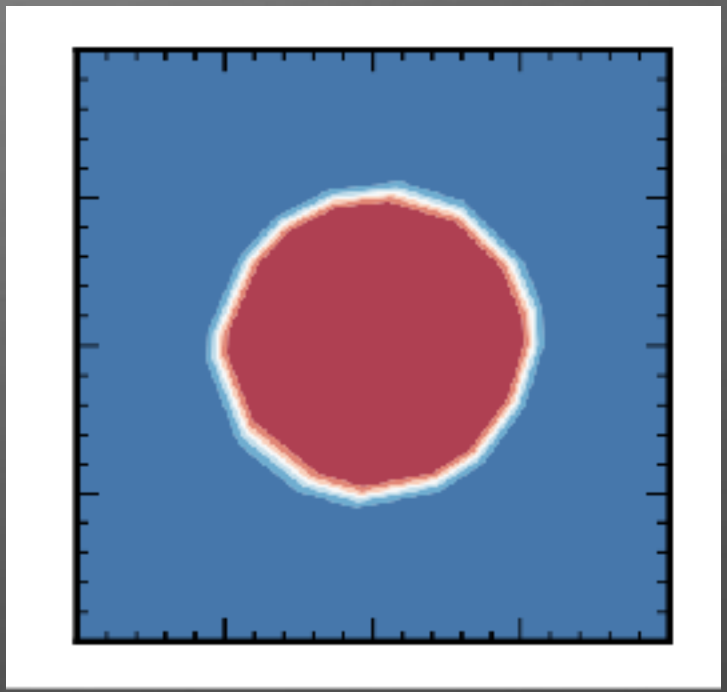
Deep Neural Network Output

Interpreting NN Classifier

1709.10106 PRD SC w/ Cohen, Ostdiek

Training data





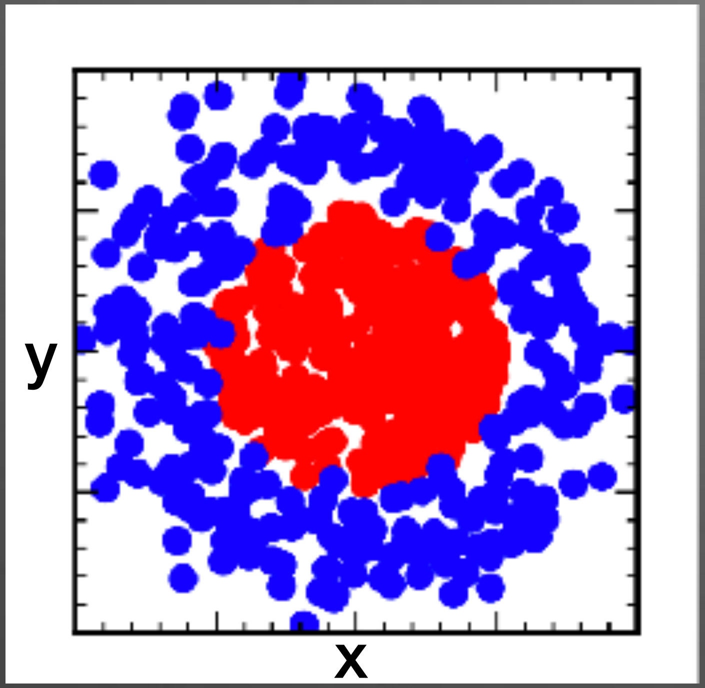
Deep Neural Network Output

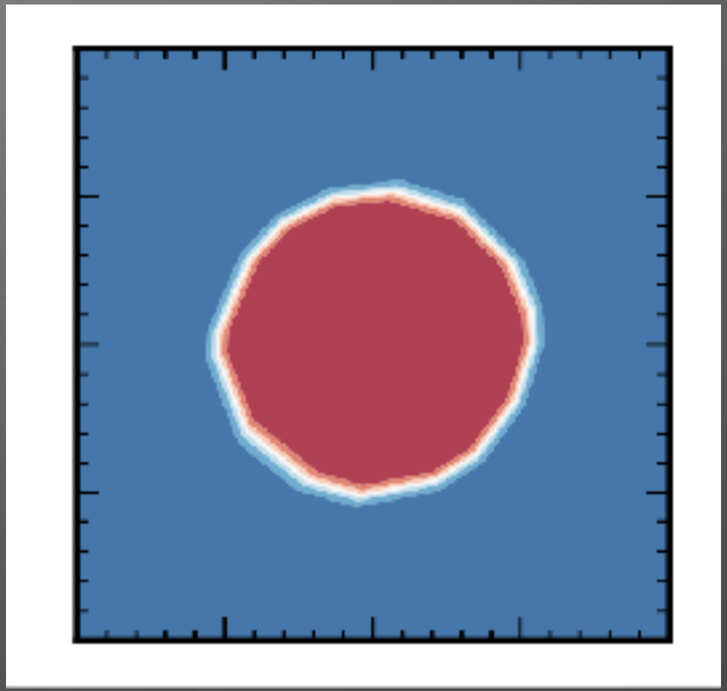
In a simple scenario can interpret NN (e.g. approximates radius)

Interpreting NN Classifier

1709.10106 PRD SC w/ Cohen, Ostdiek

Training data





Deep Neural Network Output

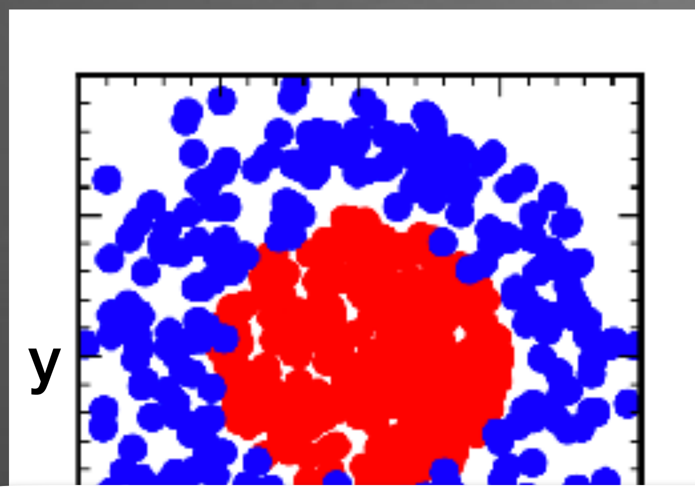
In a simple scenario
can interpret NN
(e.g. approximates radius)

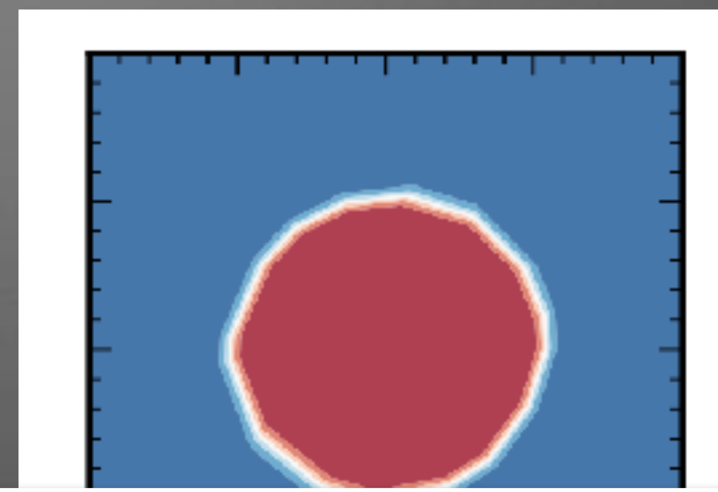
As scenario
gets complicated in input
and feature space
this is progressively
more challenging

Interpreting NN Classifier

1709.10106 PRD SC w/ Cohen, Ostdiek

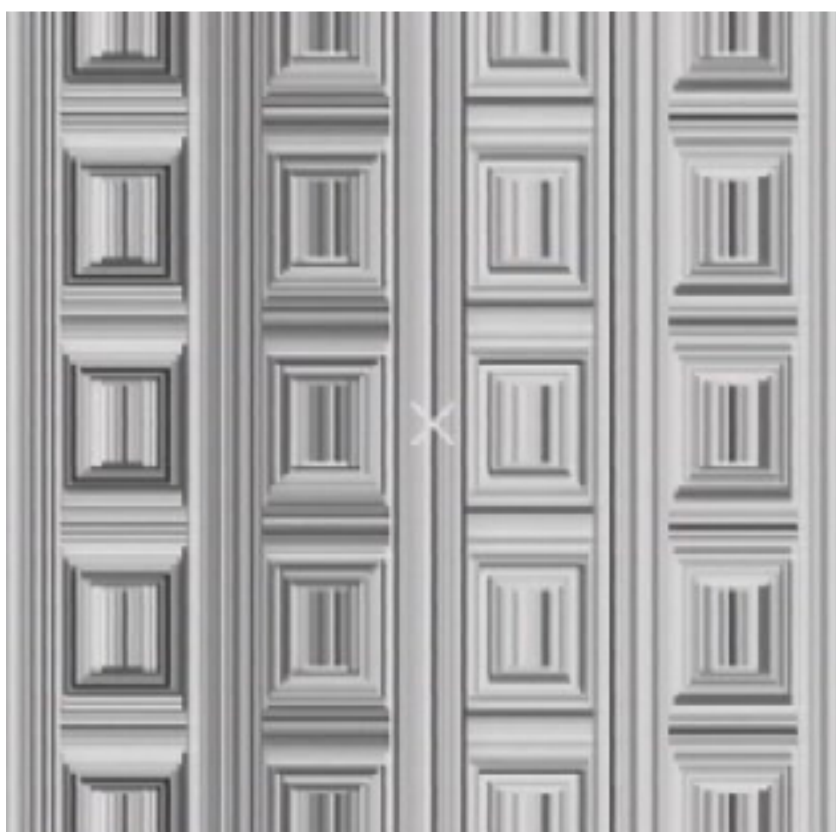
Training data





Deep Neural Network Output

In a … input
ca … ce
(e.g. a … ely
… g



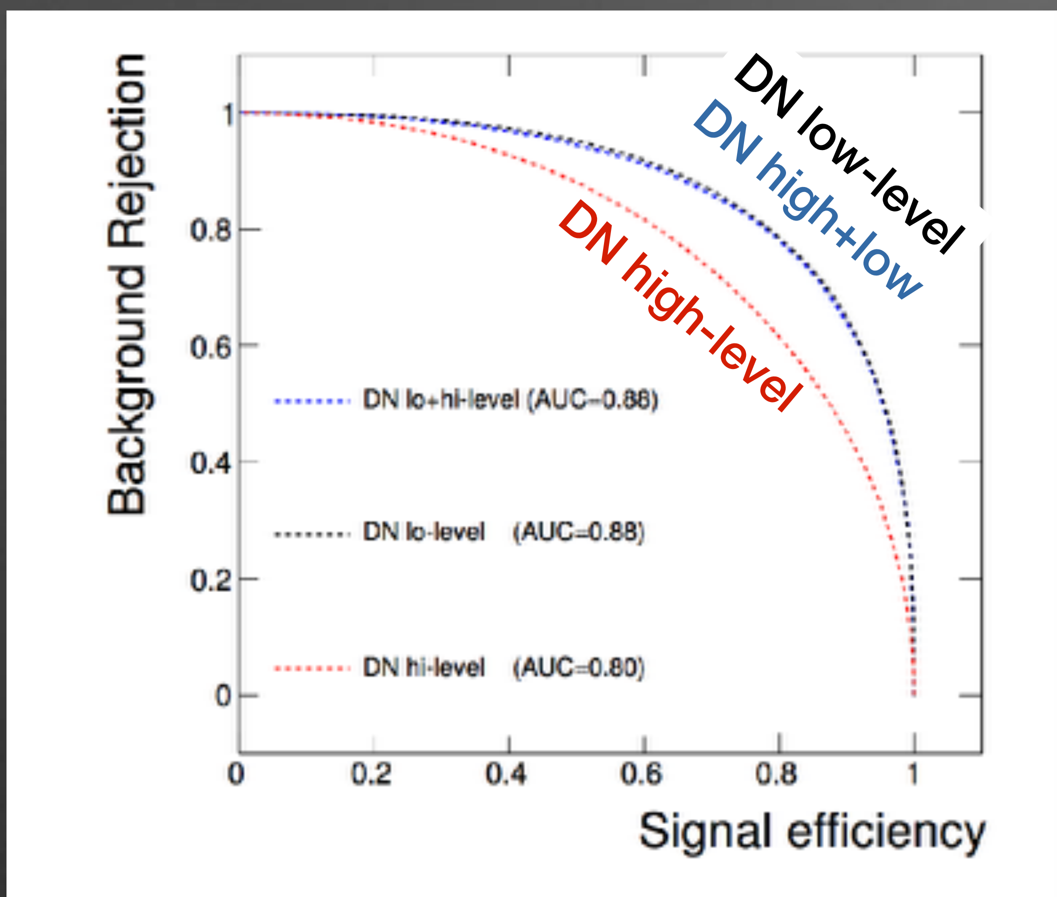
How many circles do you see?
Coffer Illusion
A. Norcia

Existing Technique: Saturation

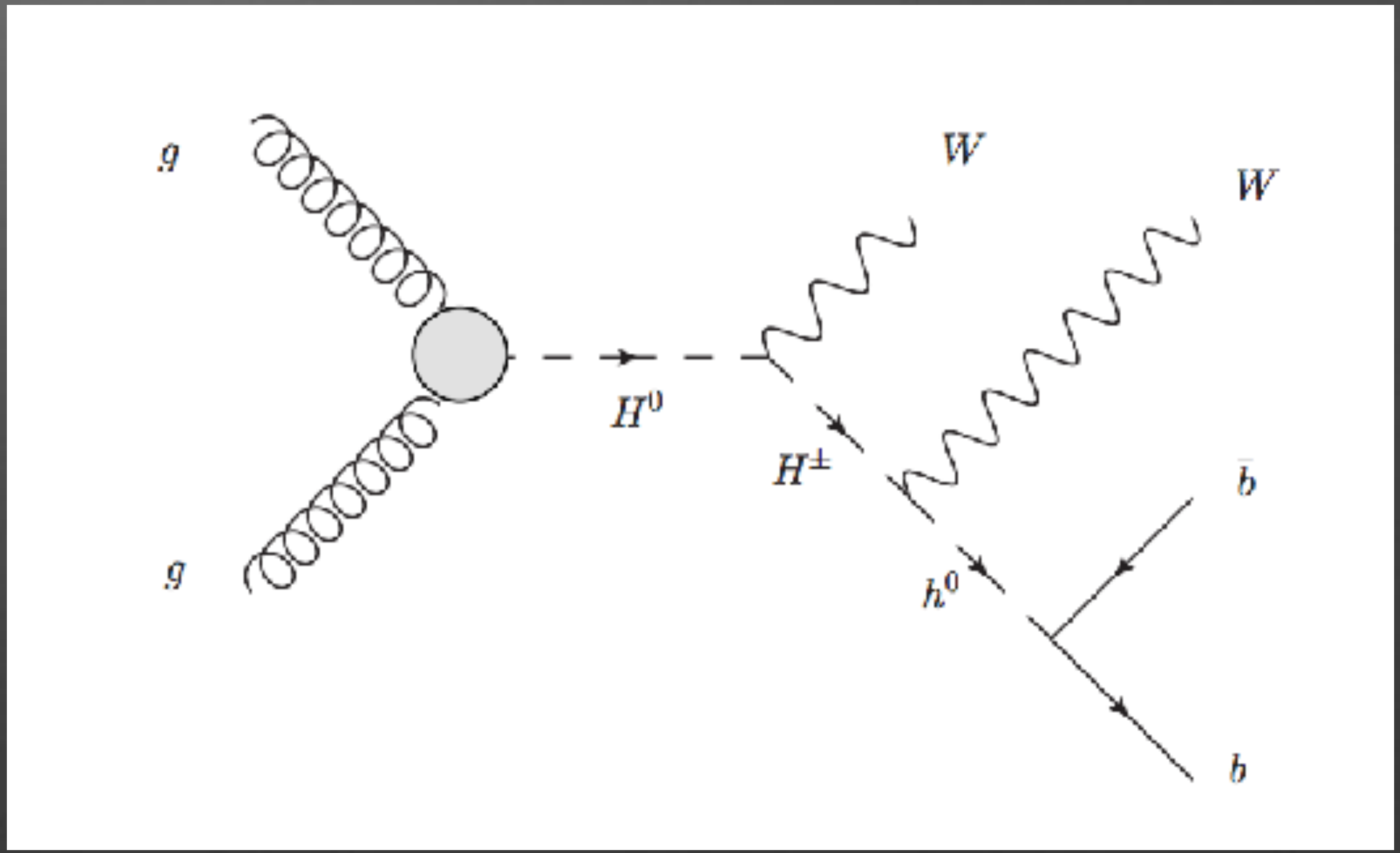
Adding a high level variable as input
and seeing if discrimination saturates
tests if classifier is sensitive to the variable

Existing Technique: Saturation

Adding a high level variable as input
and seeing if discrimination saturates
tests if classifier is sensitive to the variable

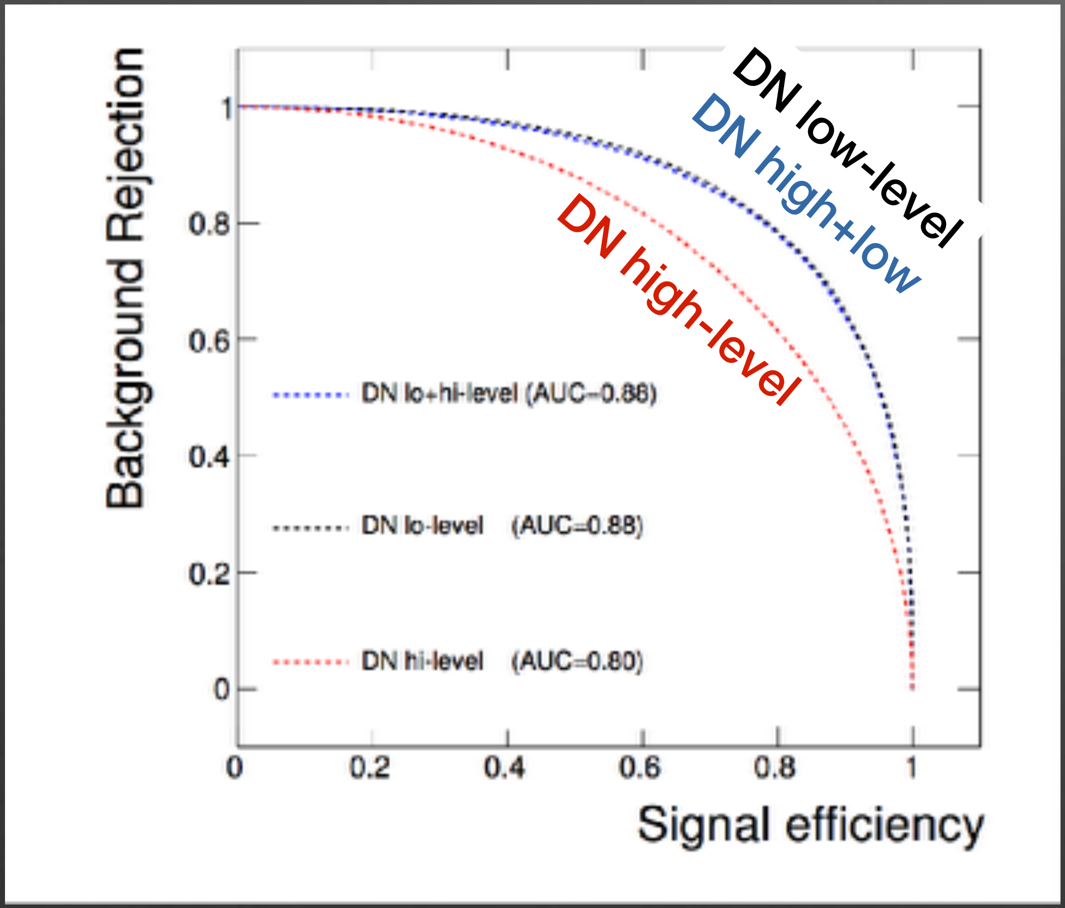


Baldi et.al. 1402.4735
High level = invariant masses
of cascade decay

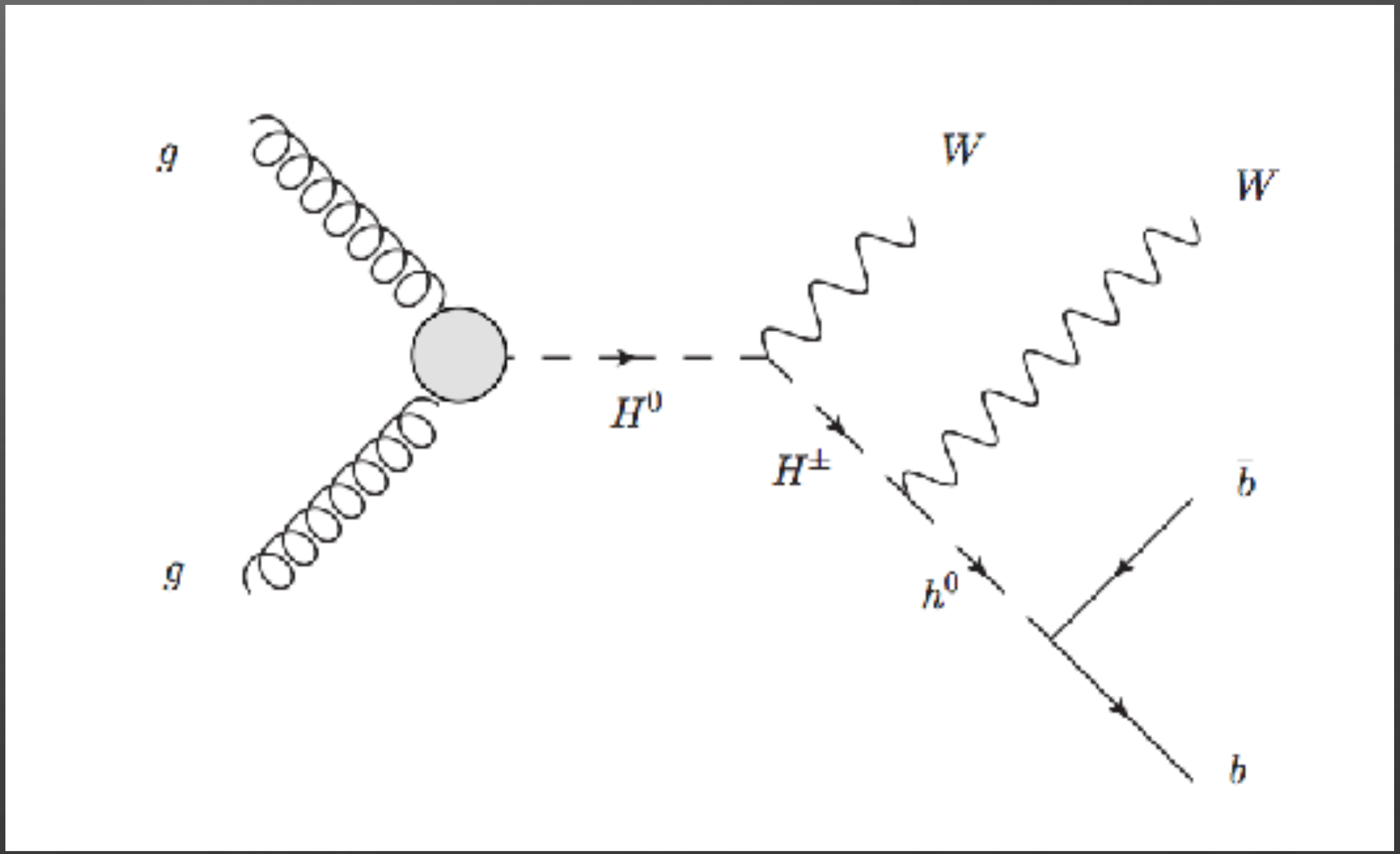


Existing Technique: Saturation

Adding a high level variable as input
and seeing if discrimination saturates
tests if classifier is sensitive to the variable



Baldi et.al. 1402.4735
High level = invariant masses
of cascade decay



So deep NN is aware of Lorentz
Invariant information without
knowing special relativity!

Data Planing

("uniform phase space" suggested in 1511.05190)

Our proposal is to remove information from events then diagnose importance through degradation of performance

Data Planing

("uniform phase space" suggested in 1511.05190)

Our proposal is to remove information from events then diagnose importance through degradation of performance

Planing involves reweighting events in a chosen variable to remove info

Data Planing

("uniform phase space" suggested in 1511.05190)

Our proposal is to remove information from events then diagnose importance through degradation of performance

Planing involves reweighting events in a chosen variable to remove info

Data Planing

("uniform phase space" suggested in 1511.05190)

Our proposal is to remove information from events then diagnose importance through degradation of performance

Planing involves reweighting events in a chosen variable to remove info

Collider Example

Signal (Z')

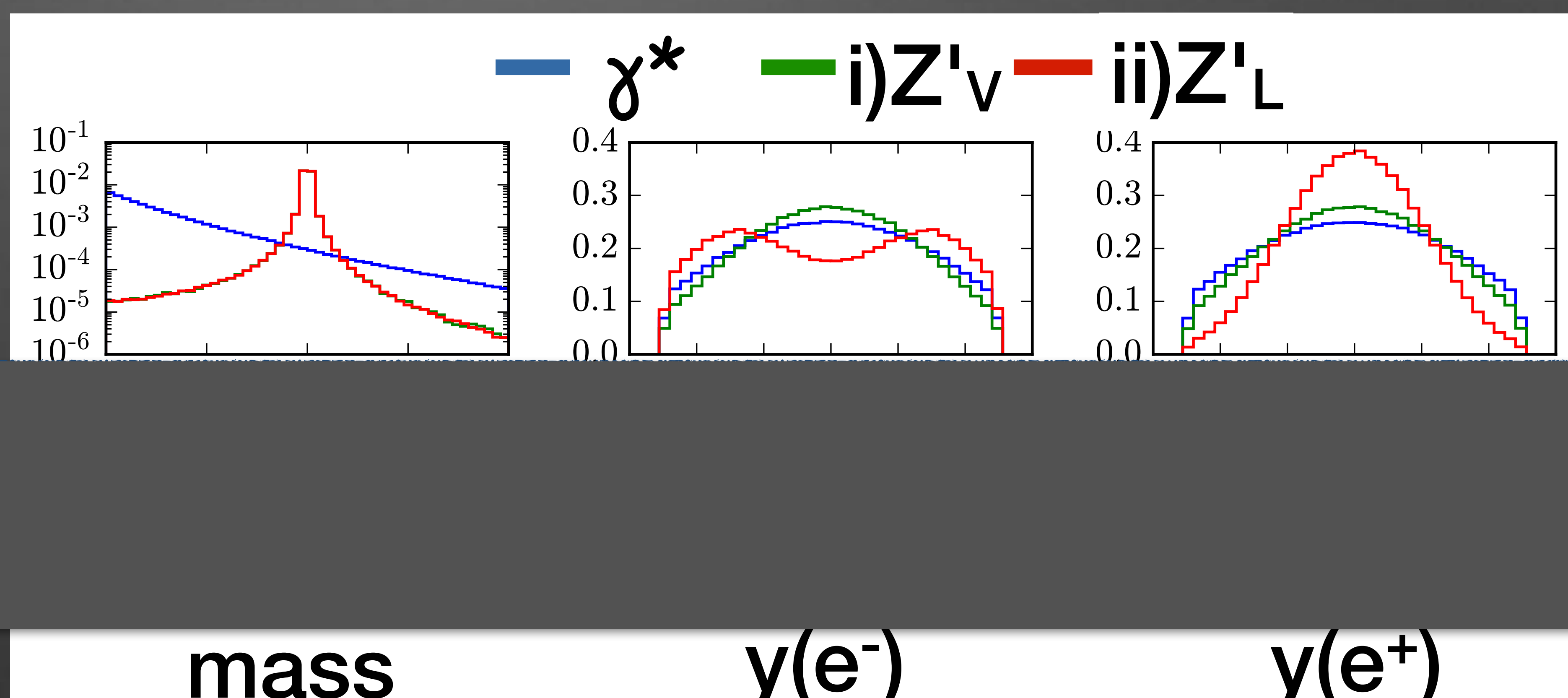
Background (photon ɣ)

Collider Example

Signal (Z')

Background (photon γ)

Raw distributions



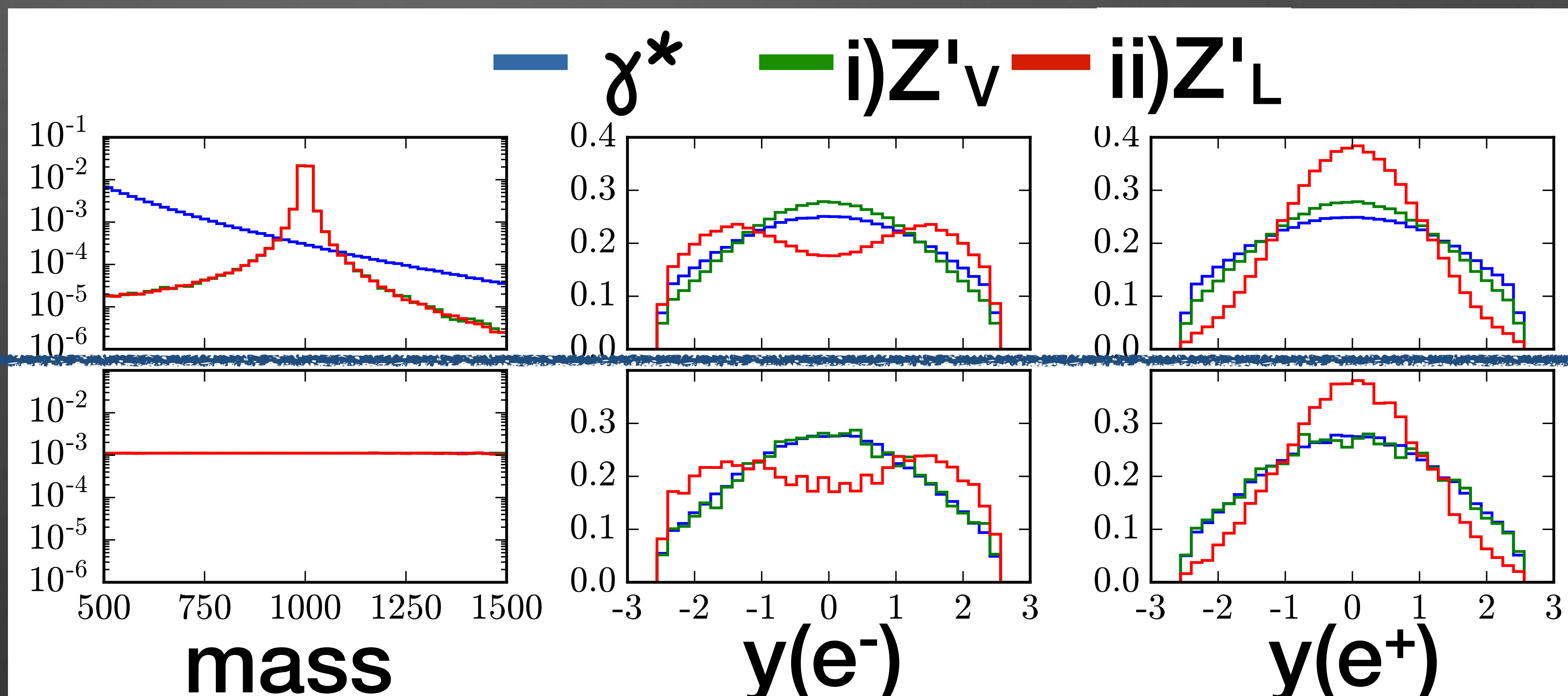
Collider Example

Signal (Z')

Background (photon γ)

Raw distributions

Planed in mass



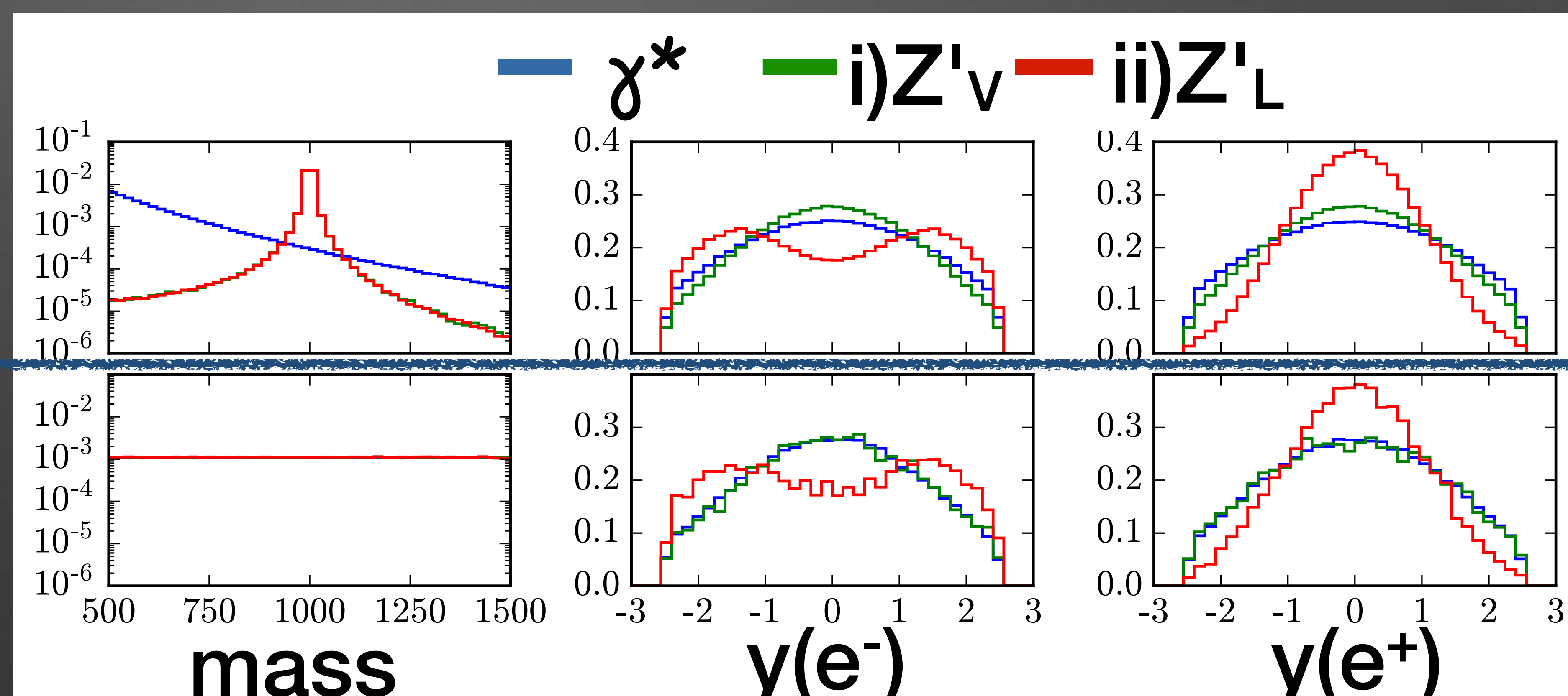
Collider Example

Signal (Z')

Background (photon γ)

Raw distributions

Planed in mass



Discriminants: Invariant mass important to both i, ii

Rapidity important for ii

Chiral Z' model

Inputs				
$(E, \vec{p})$	m	PLANED	LINEAR AUC	DEEP AUC
✓	✘	✘	$0.763280(05)$	$0.989353(59)$
✓	✓	✘	$0.942004(02)$	$0.989826(10)$
✓	✘	$\boldsymbol{m}$	$0.626648(28)$	$0.6258(24)$
✓	✘	$(\boldsymbol{m}, \boldsymbol{\Delta}\lvert\boldsymbol{y}\rvert)$	$0.52421(15)$	$0.5320(25)$

Note: AUC (Area Under Curve)
1 is perfect
0.5 is random guess

Chiral Z' model

Inputs

$(E, \vec{p})$	m	Planed	Linear AUC	Deep AUC
✓	✘	✘	0.763280(05)	0.989353(59)
✓	✓	✘	0.942004(02)	0.989826(10)
✓	✘	$\boldsymbol{m}$	0.626648(28)	0.6258(24)
✓	✘	$(\boldsymbol{m}, \boldsymbol{\Delta}\vert\boldsymbol{y}\vert)$	0.52421(15)	0.5320(25)

Note: AUC (Area Under Curve)
1 is perfect
0.5 is random guess

Saturation: Deep NN aware of mass

Chiral Z' model

Inputs

$(E, \vec{p})$	m	Planed	Linear AUC	Deep AUC
✓	✘	✘	0.763280(05)	0.989353(59)
✓	✓	✘	0.942004(02)	0.989826(10)
✓	✘	m	0.626648(28)	0.6258(24)
✓	✘	$(m, \Delta\|y\|)$	0.52421(15)	0.5320(25)

Note: AUC
(Area Under Curve)
1 is perfect
0.5 is random
guess

Planing in mass removes most
of distinguishing power, leaving linear info?

Chiral Z' model

Inputs				
$(E, \vec{p})$	m	Planed	Linear AUC	Deep AUC
✓	✘	✘	0.763280(05)	0.989353(59)
✓	✓	✘	0.942004(02)	0.989826(10)
✓	✘	m	0.626648(28)	0.6258(24)
✓	✘	$(m, \Delta\|y\|)$	0.52421(15)	0.5320(25)

Note: AUC
(Area Under Curve)
1 is perfect
0.5 is random guess

Plane in mass and Δ|y| removes essentially all info

Pros

- Large dynamic range diagnostic
- No need to change architecture

Pros

- Large dynamic range diagnostic
- No need to change architecture

Cons

- Scalable to multiple variables?
- Systematic to explore or needs physics intuition?

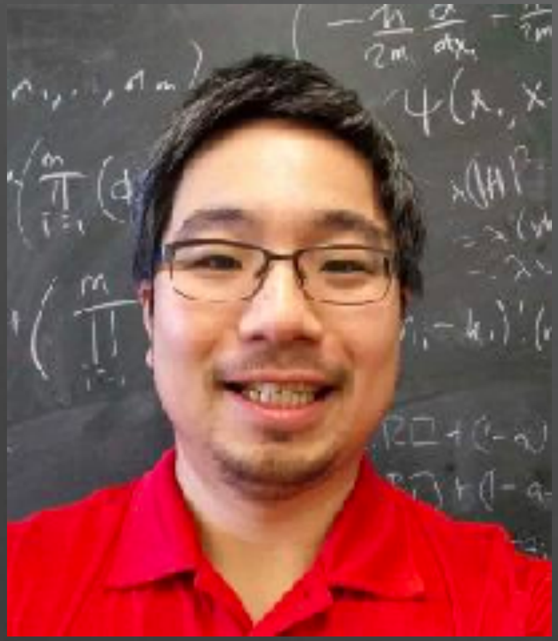
Future Directions

- Apply to nontrivial scenarios (e.g. jet substructure, check saturation)
- Can one discover a new useful variable?
- Can one plane in output of a NN to test if discrimination was optimized?

Conclusions

- Planing is a useful diagnostic tool to understand machine learning classifiers
- Weighting to remove info about a variable allows degradation to test its importance
- Machine learning is fascinating and basic questions are still unanswered

Thanks for your attention!



chang2@uoregon.edu

Extra Slides

NNs are good at classification because a large enough NN can approximate any function (Universality Theorem)

1-D Physics-level proof

NNs are good at classification because a large enough NN can approximate any function (Universality Theorem)

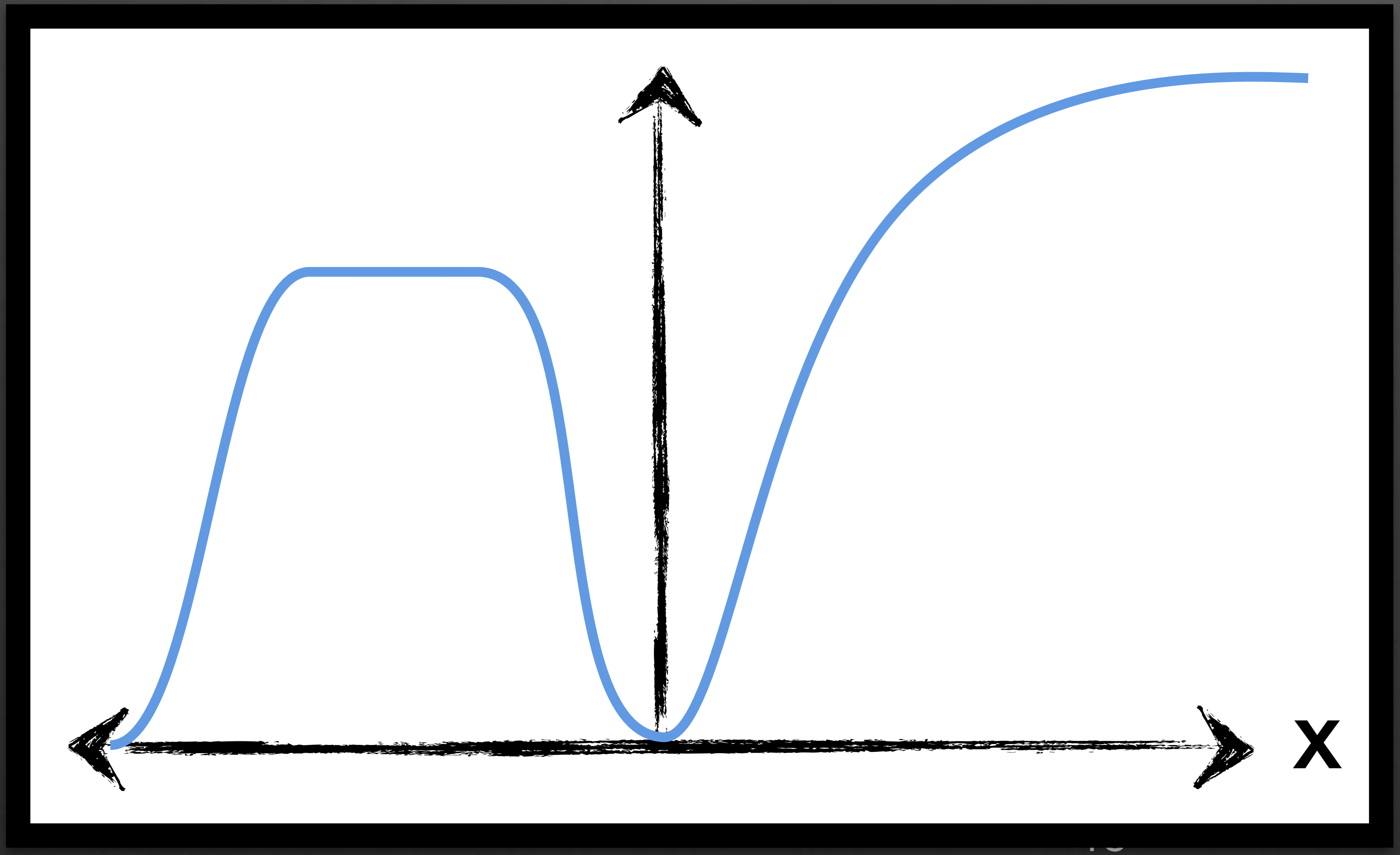
1-D Physics-level proof

$$\frac{1}{1+\exp\left[-w(x-x_0)\right]} \xrightarrow{w\to\infty} \theta(x-x_0)$$

NNs are good at classification because a large enough NN can approximate any function (Universality Theorem)

1-D Physics-level proof

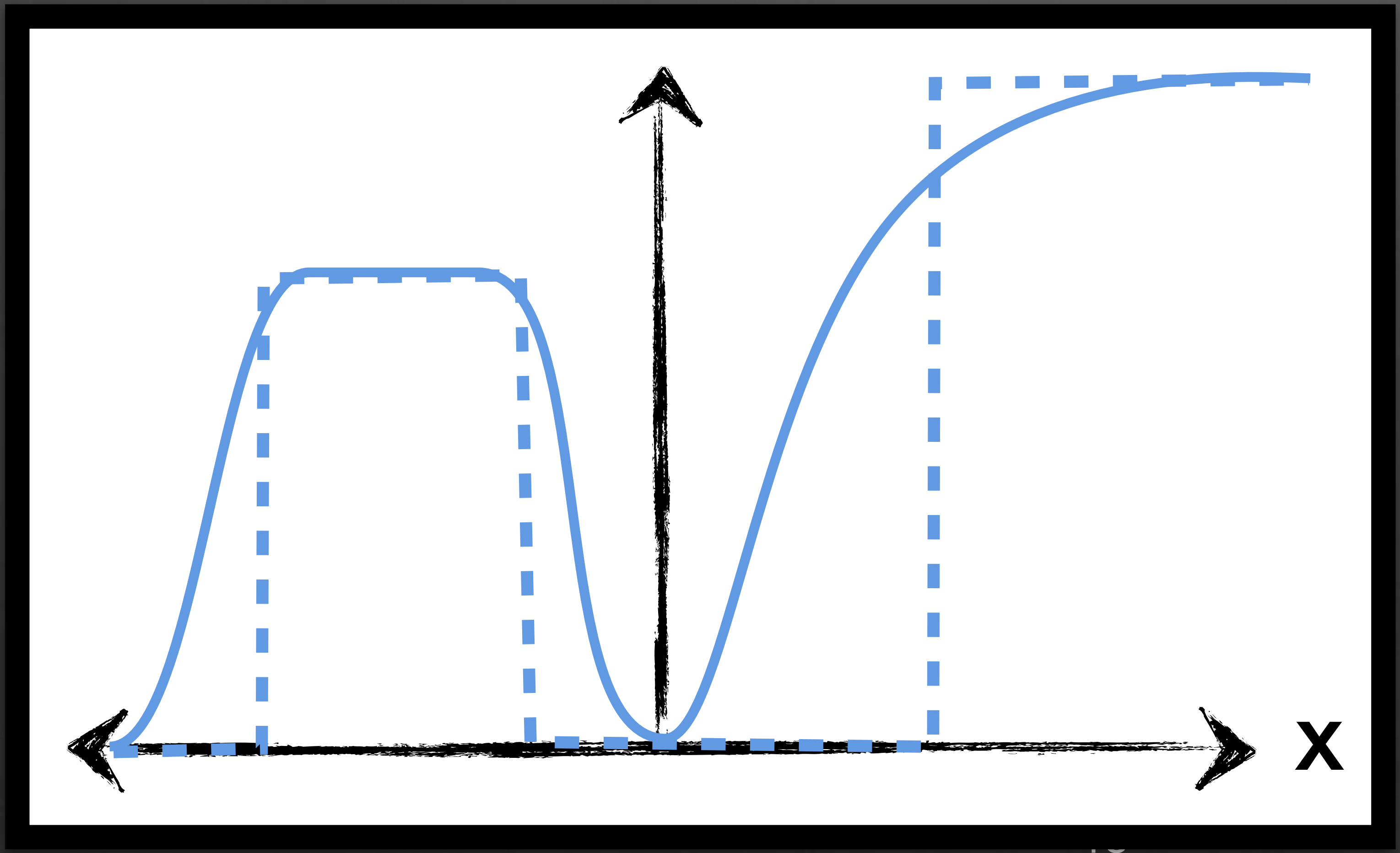
$$\frac{1}{1+\exp\left[-w(x-x_0)\right]} \xrightarrow{w\to\infty} \theta(x-x_0)$$



NNs are good at classification because a large enough NN can approximate any function (Universality Theorem)

1-D Physics-level proof

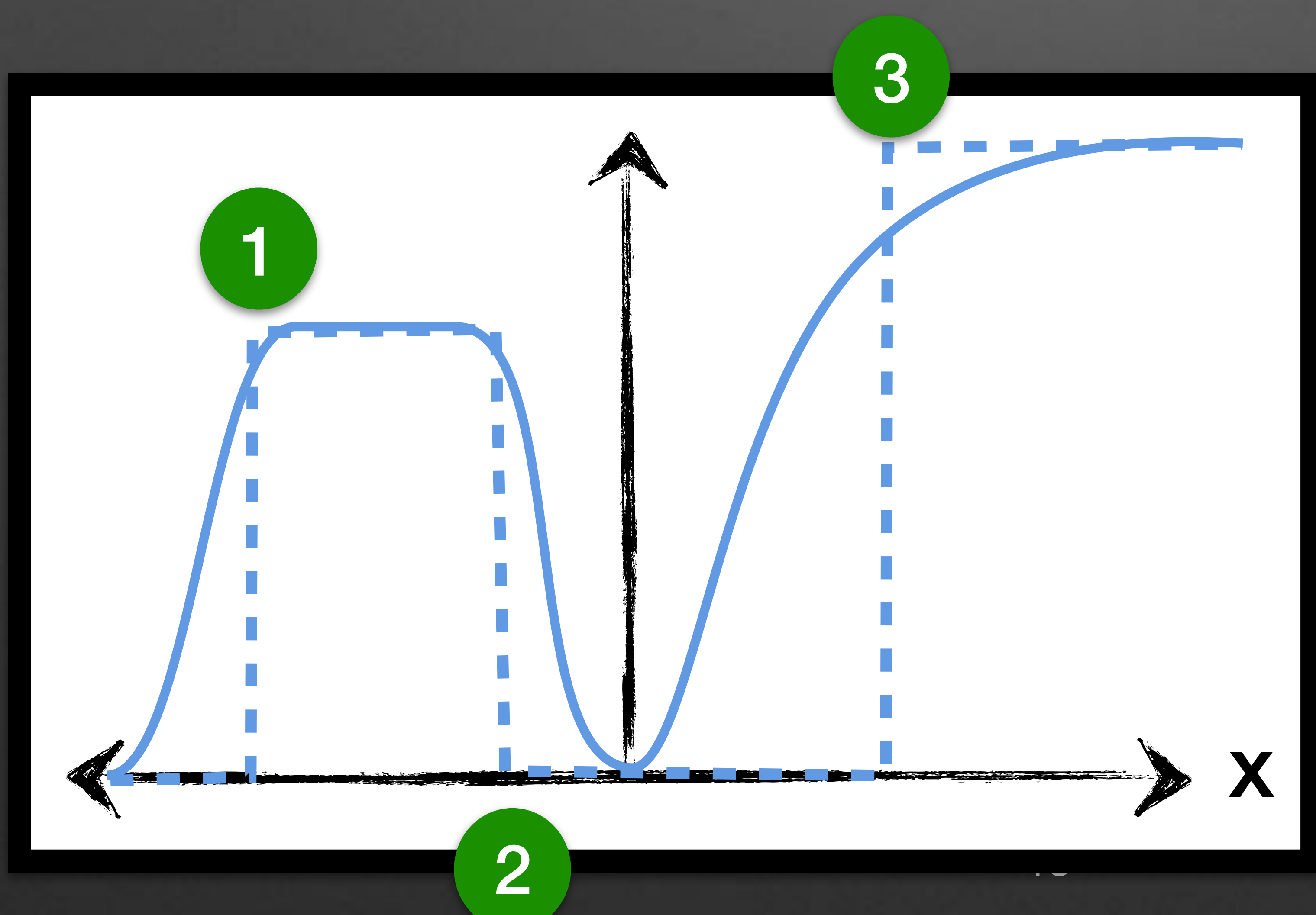
$$\frac{1}{1+\exp\left[-w(x-x_0)\right]} \xrightarrow{w\to\infty} \theta(x-x_0)$$

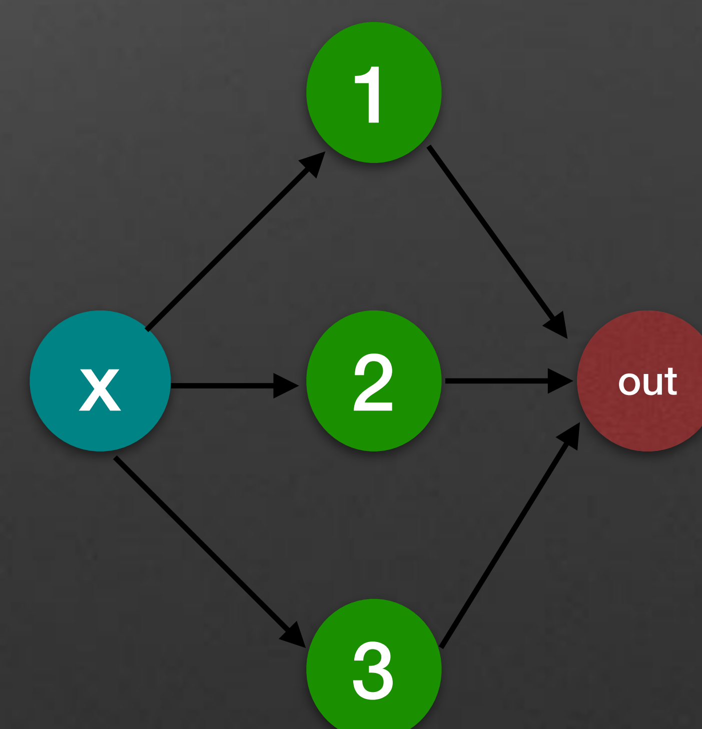


NNs are good at classification because a large enough NN can approximate any function (Universality Theorem)

1-D Physics-level proof

$$\frac{1}{1+\exp\left[-w(x-x_0)\right]} \xrightarrow{w\to\infty} \theta(x-x_0)$$

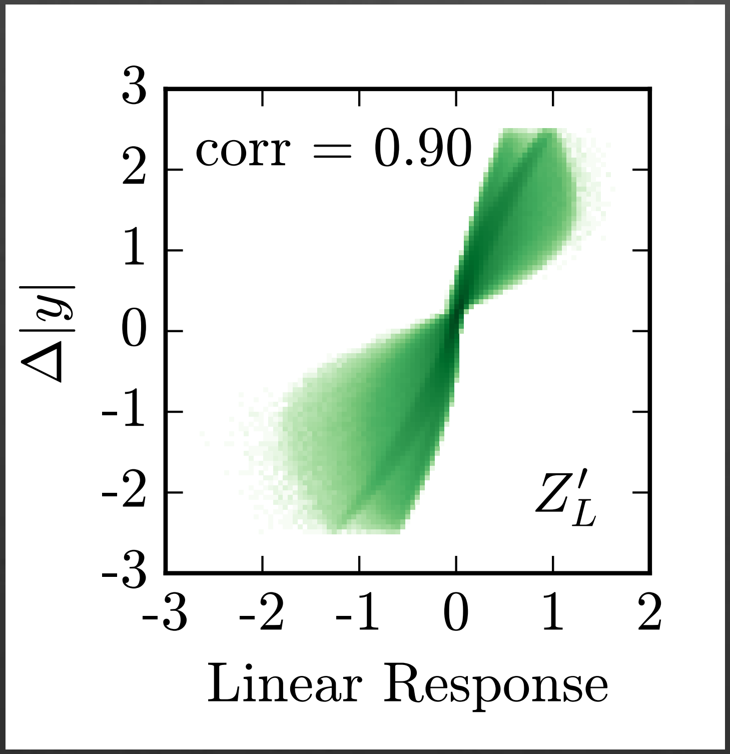




More on Chiral Z'

Inputs				
$(E, \vec{p})$	m	Planed	Linear AUC	Deep AUC
✓	✘	✘	0.763280(05)	0.989353(59)
✓	✓	✘	0.942004(02)	0.989826(10)
✓	✘	$\boldsymbol{m}$	0.626648(28)	0.6258(24)
✓	✘	$(\boldsymbol{m}, \boldsymbol{\Delta}\lvert\boldsymbol{y}\rvert)$	0.52421(15)	0.5320(25)

Note: AUC
(Area Under Curve)
1 is perfect
0.5 is random
guess



Rapidity difference
close to linear

Vector Z' Model

$(E, \vec{p})$	m	Planed	Linear AUC	Deep AUC
✓	✗	✗	0.746221(01)	0.988510(98)
✓	✓	✗	0.938967(01)	0.989007(03)
✓	✗	$\boldsymbol{m}$	0.50550(29)	0.4942(48)

TABLE II: The AUC output for a variety of input configurations applied to the Z'_V model and the photon background.

Baldi et.al.

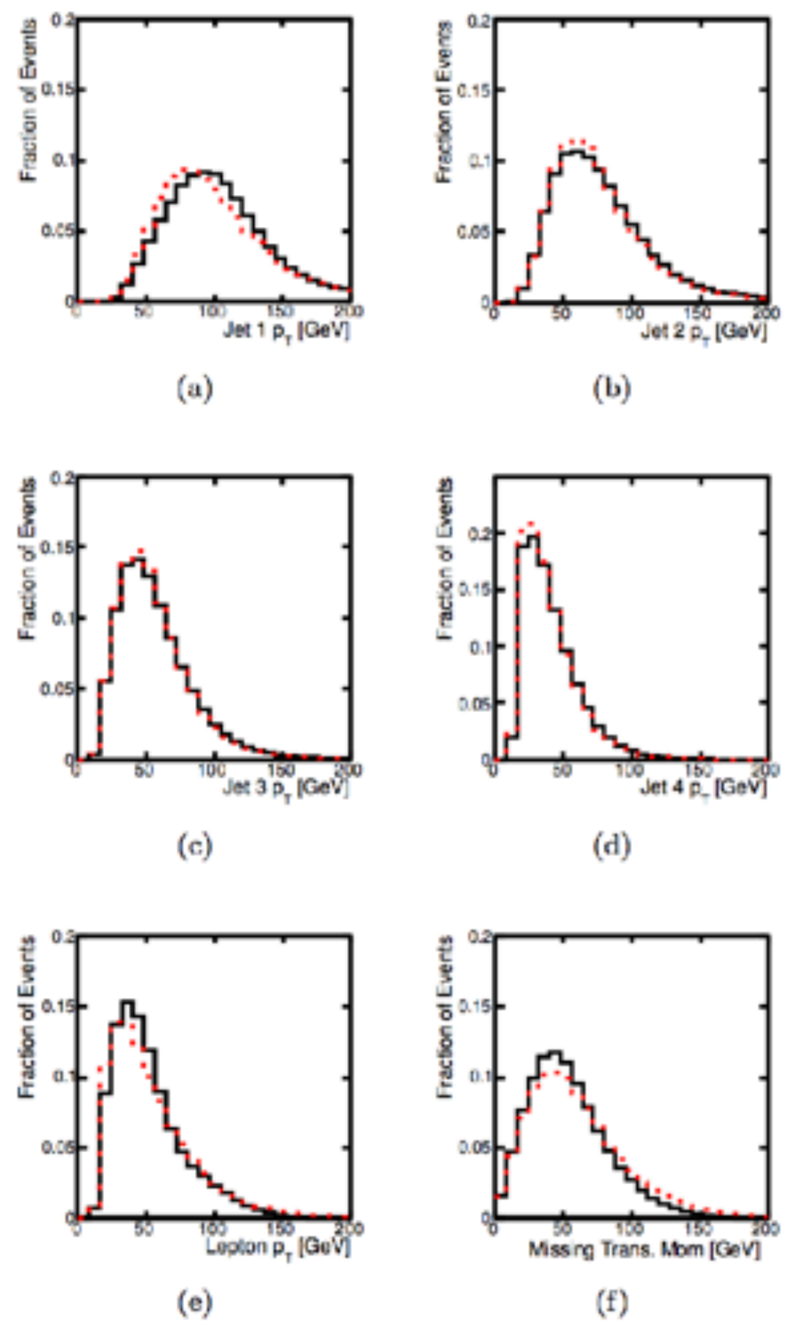
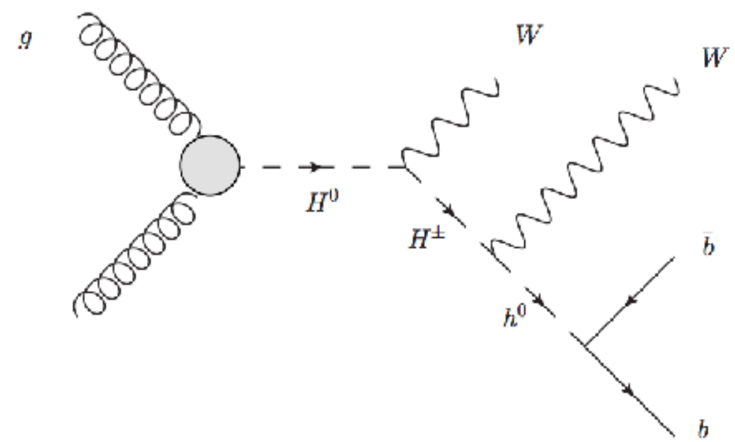


FIG. 2: **Low-level input features for Higgs benchmark.** Distributions in $\ell\nu jjb\bar{b}$ events for simulated signal (black) and background (red) benchmark events. Shown are the distributions of transverse momenta (p_T) of each observed particle (a,b,c,d,e) as well as the imbalance of momentum in the final state (f). Momentum angular information for each observed particle is also available to the network, but is not shown, as the one-dimensional projections have little information.



Semileptonic Cascade

Bkgd: ttbar

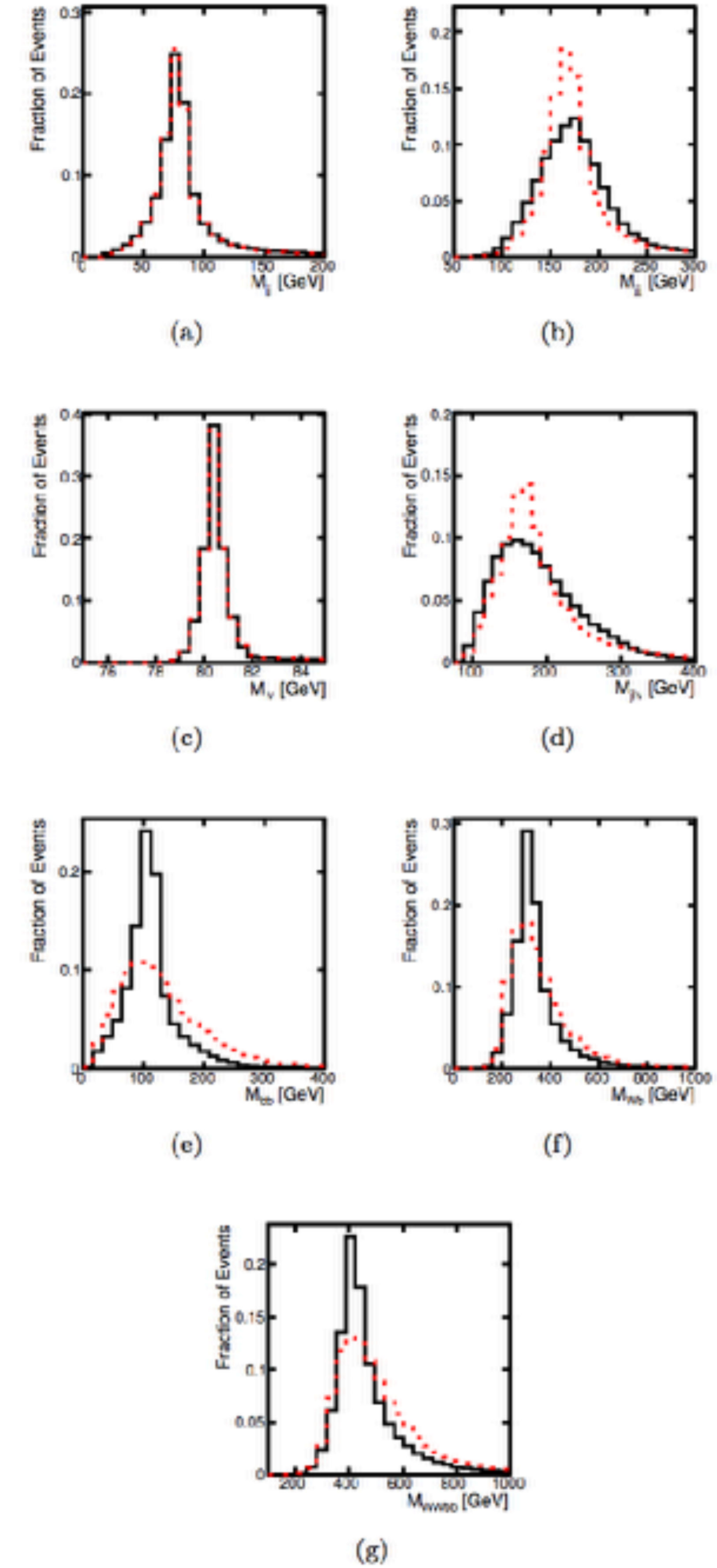


FIG. 3: **High-level input features for Higgs benchmark.** Distributions in simulation of invariant mass calculations in $\ell\nu jjb\bar{b}$ events for simulated signal (black) and background (red) events.

Other Attempts at Understanding NNs

Almeida et.al. 1501.05968 Top quark selections based on calorimeter images

Explored
i) most activating images
ii)Correlations

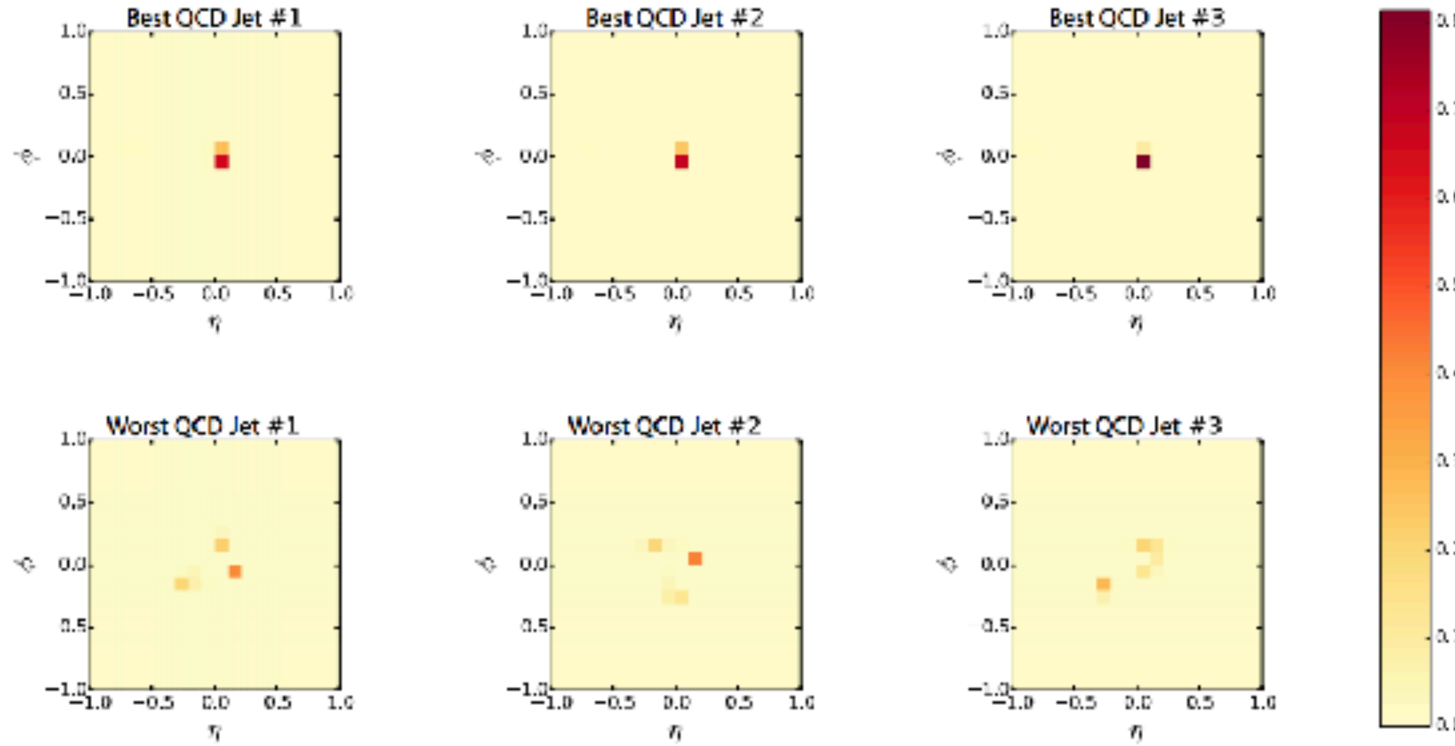


Figure 6. Energy deposit patterns for three jets with the lowest (top row) and highest (bottom row) ANN scores in the QCD jet sample with $p_T \in [800, 900]$ GeV.

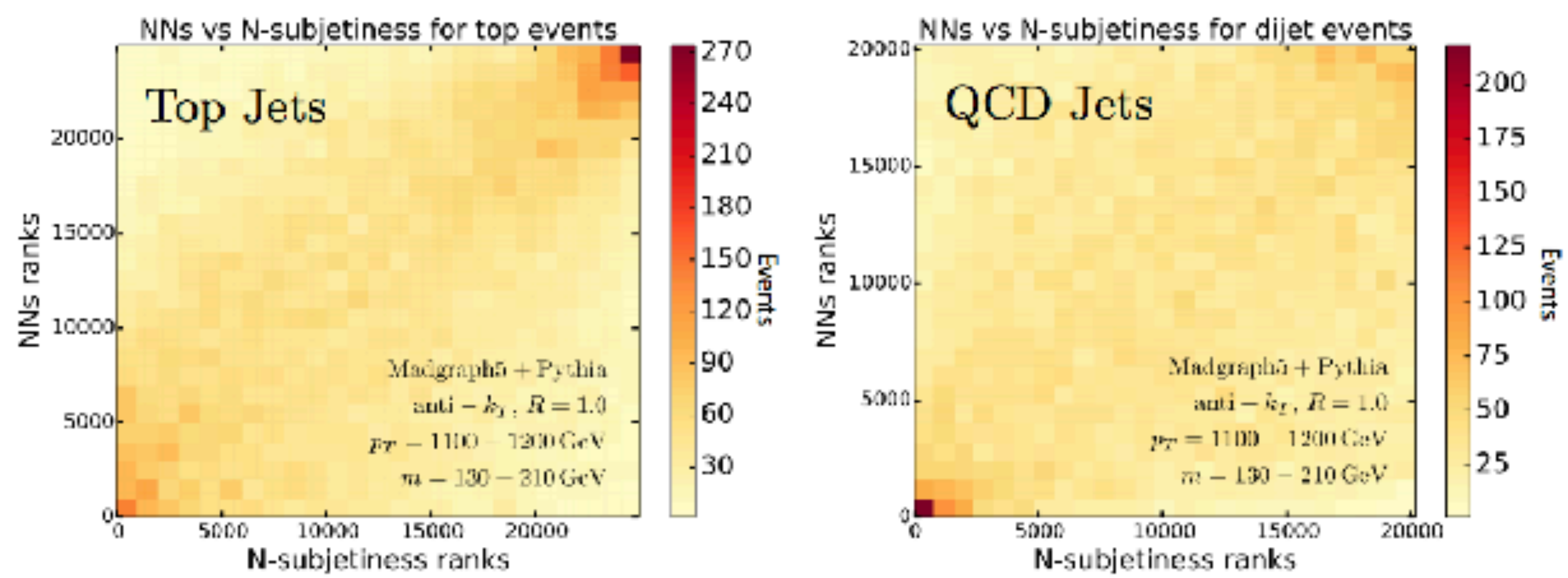
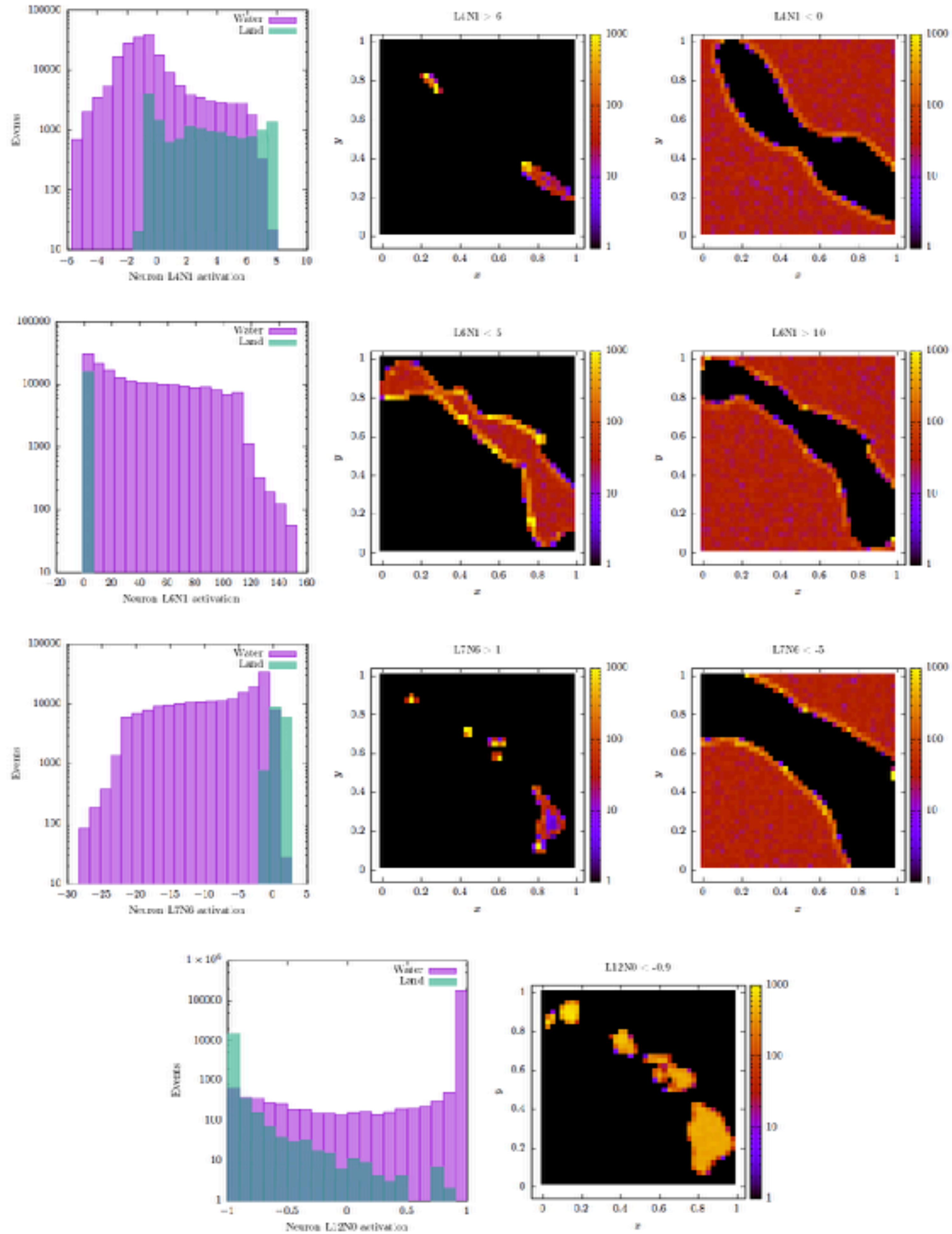


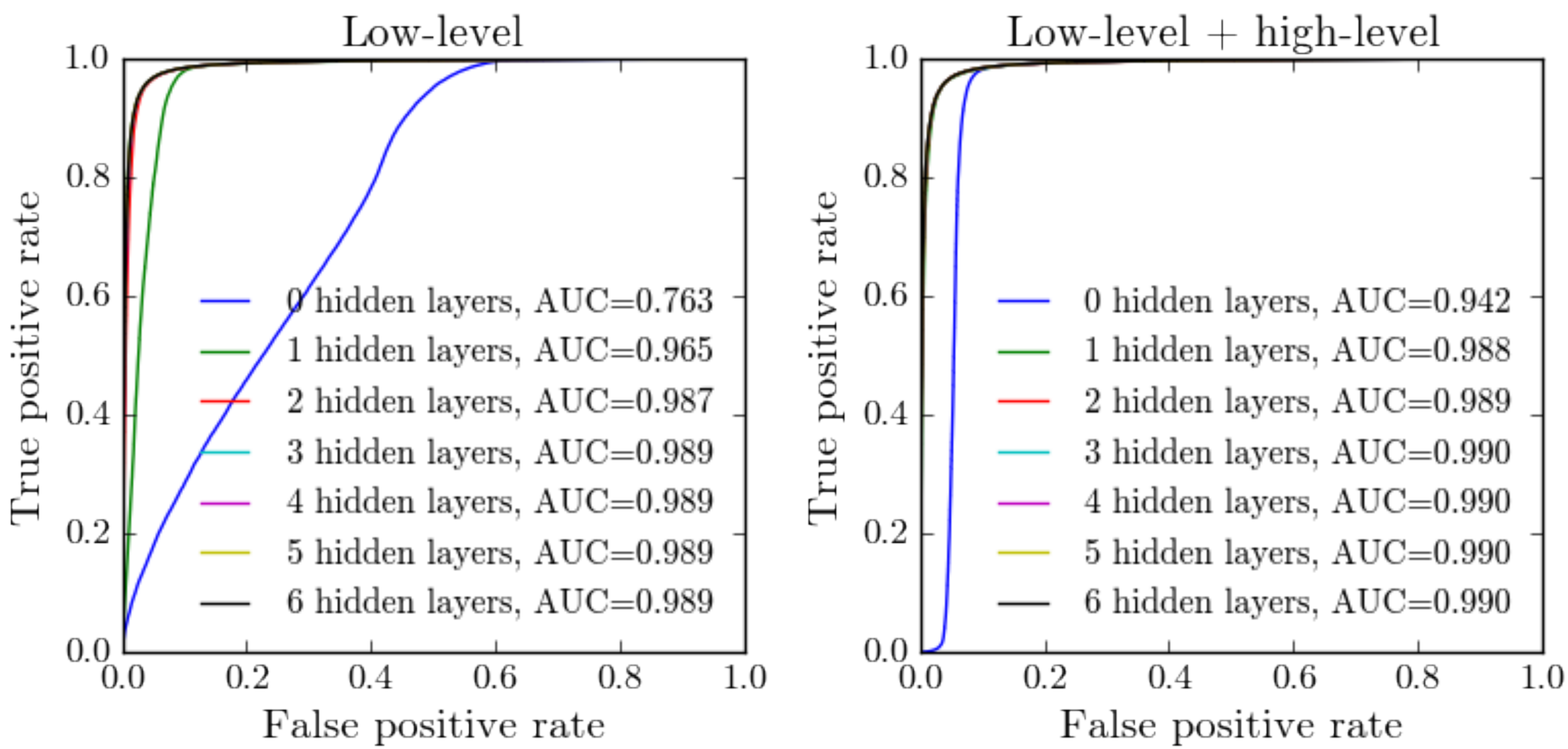
Figure 7. Correlation between the rankings of jets according to N-subjettiness (horizontal axis) and ANN score (vertical axis). Left: top sample, $p_T \in [1100, 1200]$ GeV. Right: QCD jet sample, same p_T range. Jets are ranked in order of increasing "topness" for both samples.

Roxlo, Reece
1804.09278

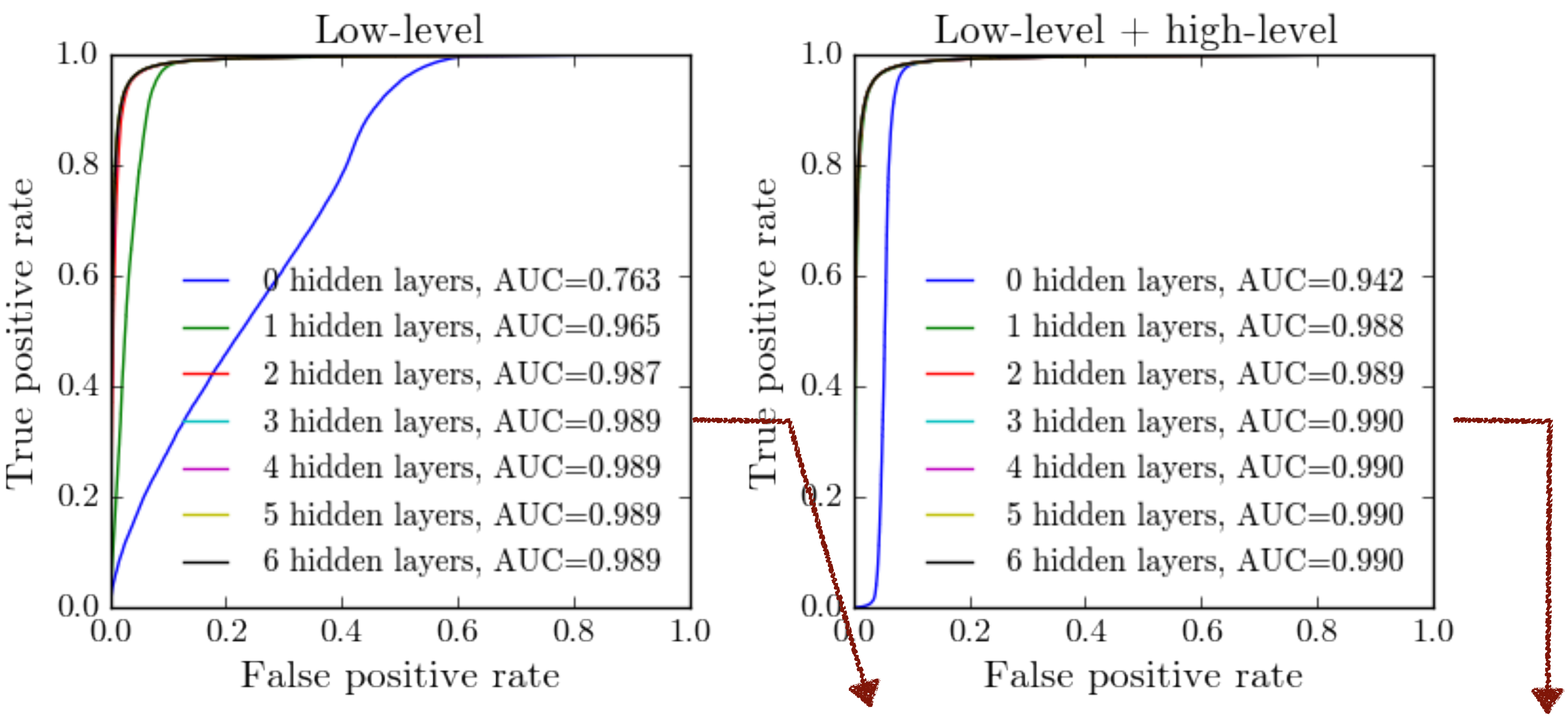
Look at late nodes of NN
Maximize Activation to see where decision boundary is



Using saturation to pick the network

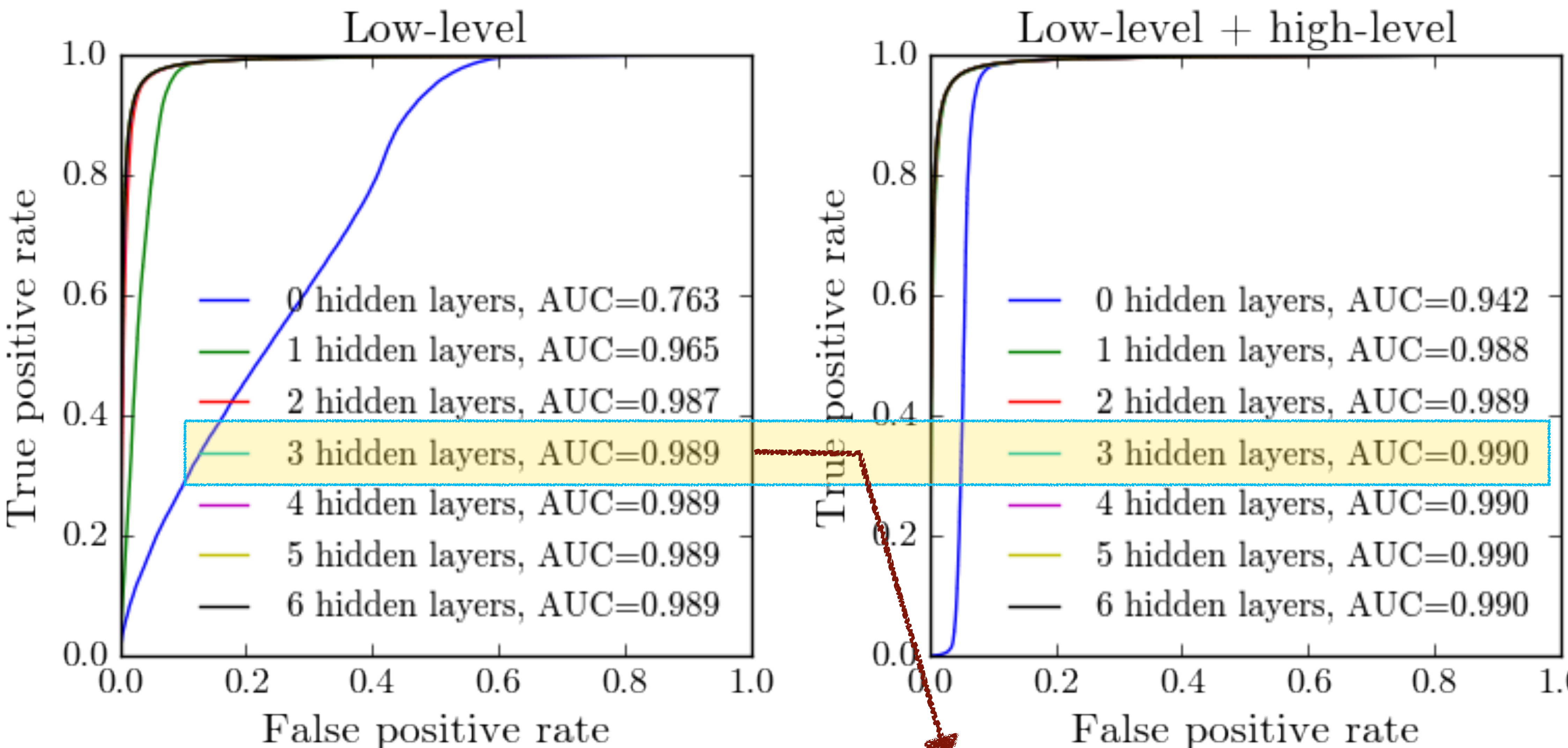


Using saturation to pick the network



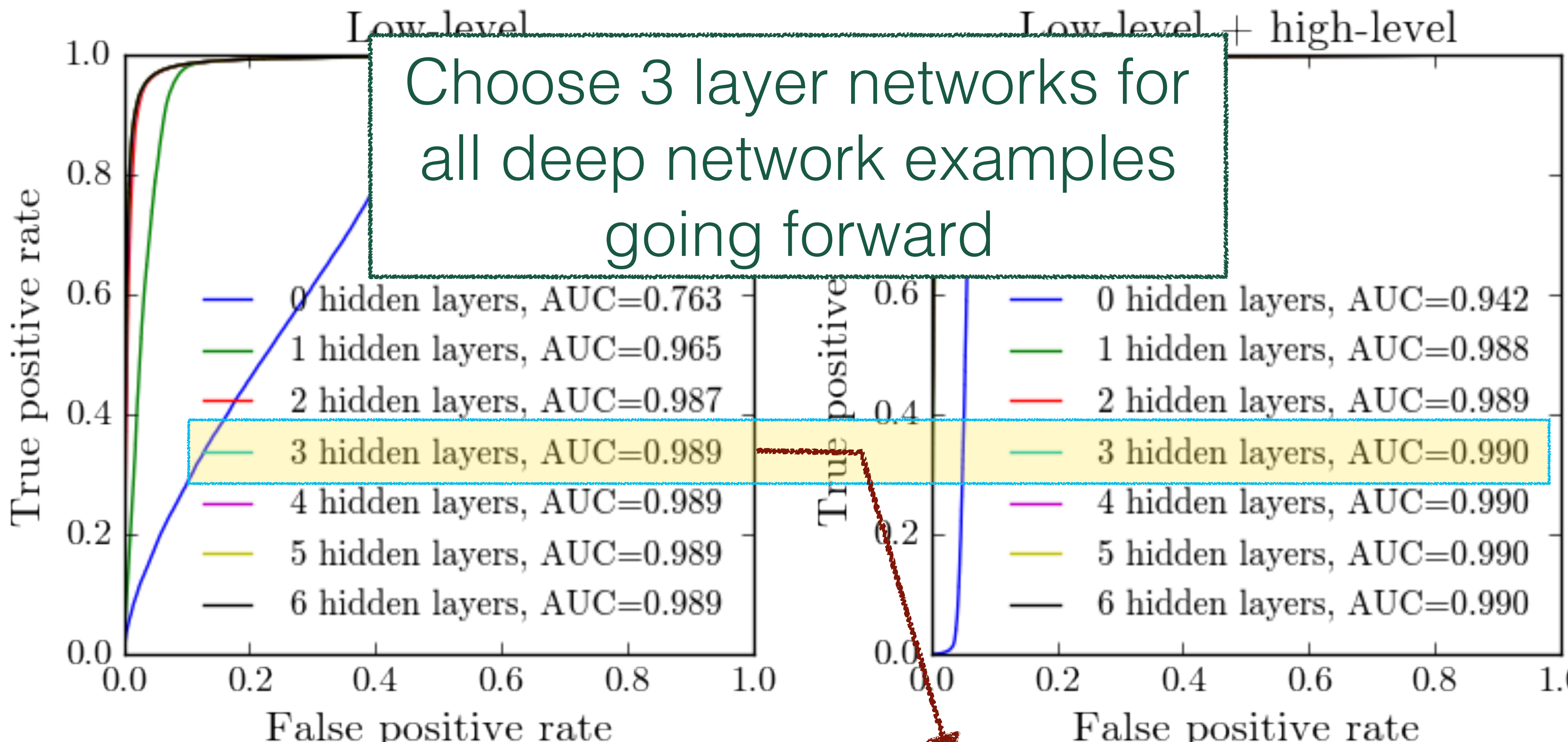
Minimal improvement

Using saturation to pick the network



high-level not adding much Minimal improvement

Using saturation to pick the network



high-level not adding much Minimal improvement