UNIVERSITY OF MARYLAND

JOHNS HOPKINS UNIVERSITY

UNIVERSITY OF MICHIGAN

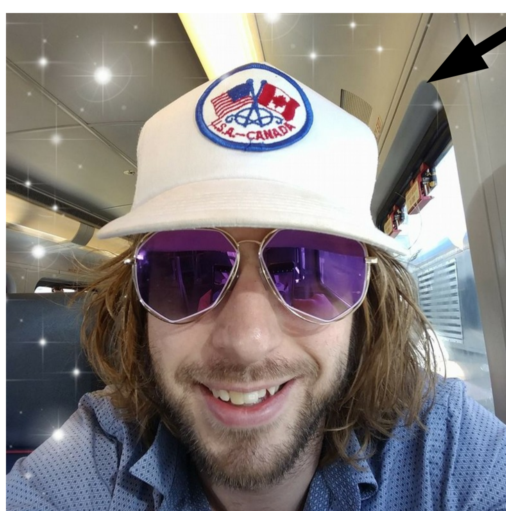
CWoLa Hunting:

Extending the Bump Hunt with Machine Learning

Based on:

Phys. Rev. Lett. 121, 241803 (2018)

[1805.02664] Jack Collins, Kiel Howe, Ben Nachman



Outline

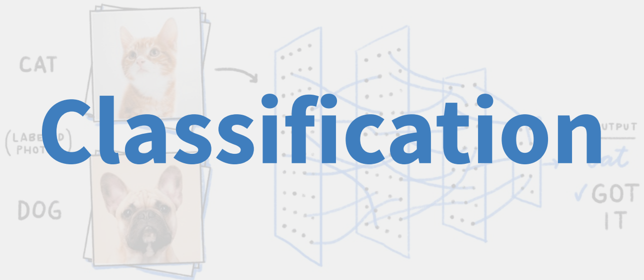
1) Machine Learning

2) Model Unspecific Searches

3) CWoLa Hunting

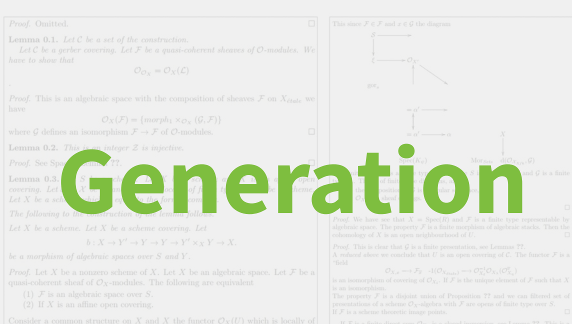
Machine Learning

Classification

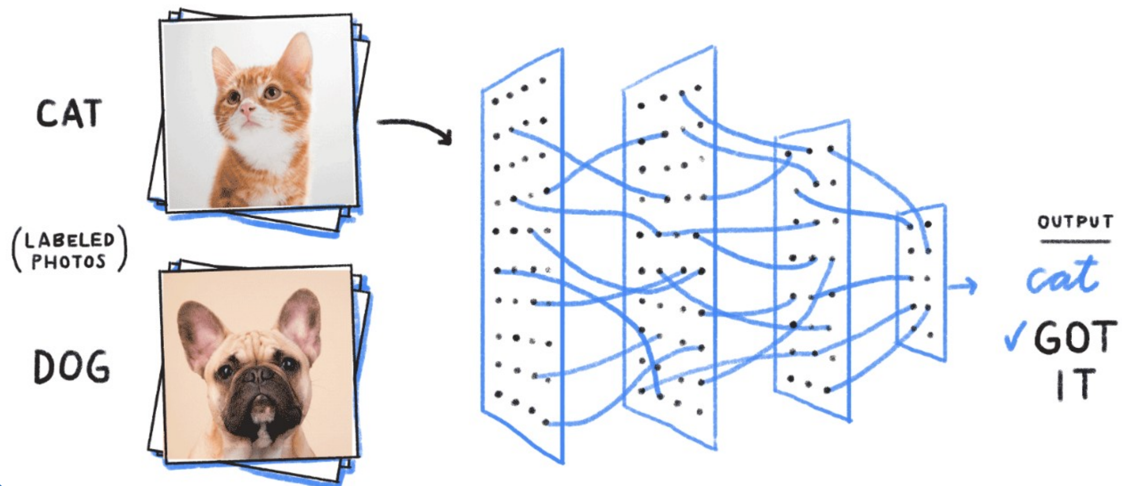


Regression

Generation



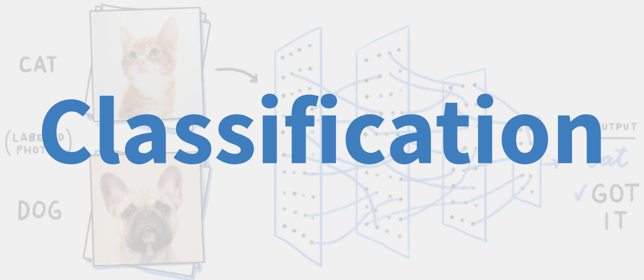
Classification



https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

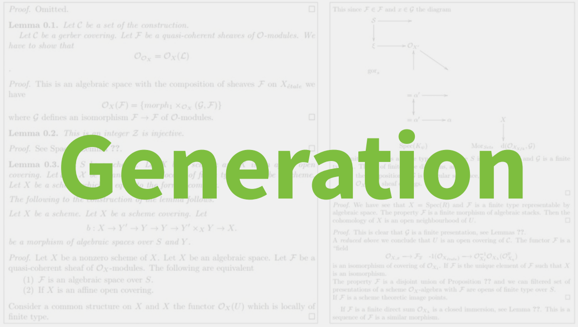
Machine Learning

Classification



Regression

Generation

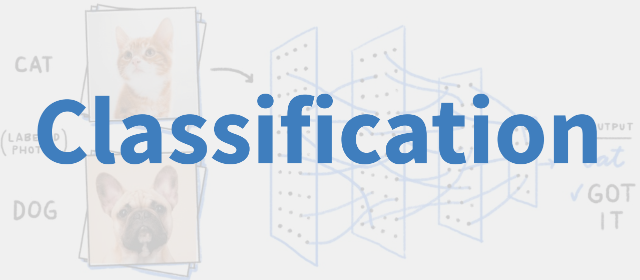


Regression

https://research.nvidia.com/sites/default/files/publications/dnn_denoise_author.pdf

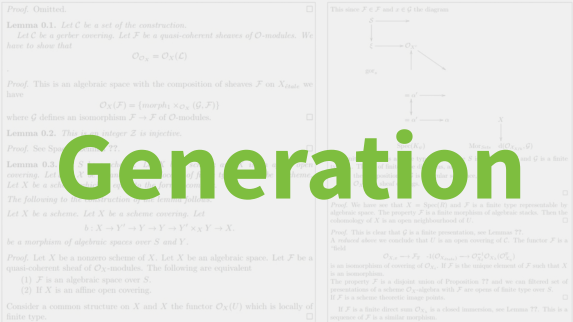
Machine Learning

Classification



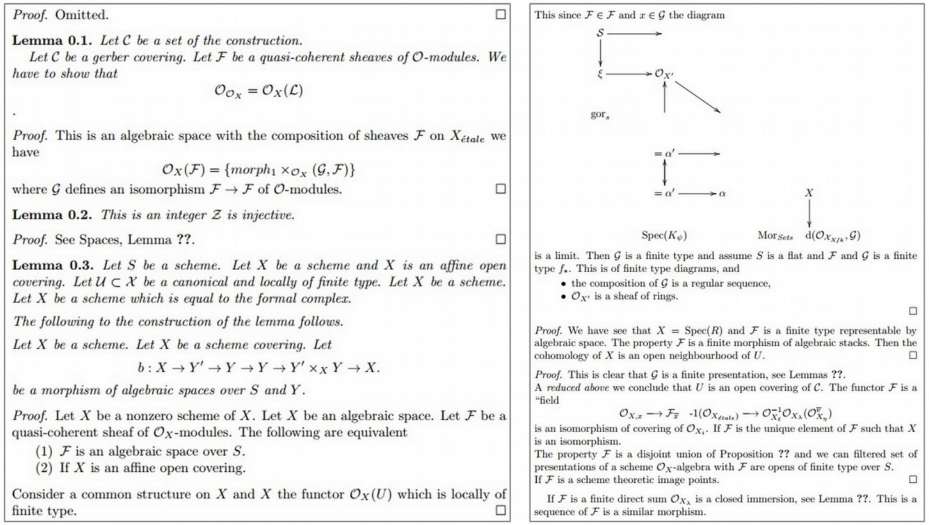
Regression

Generation



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Generation



Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{étale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram

is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \ \text{-1}(\mathcal{O}_{X_{étale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

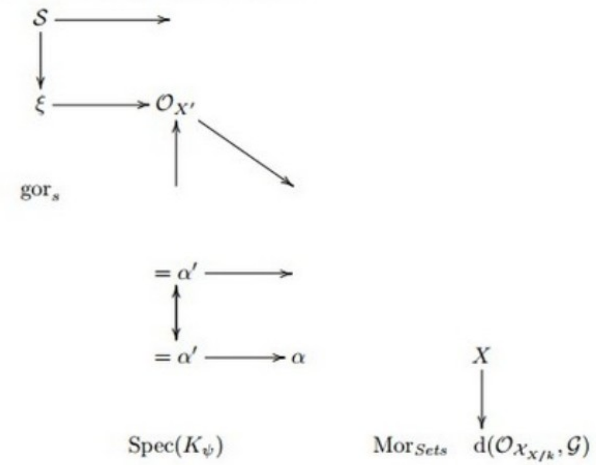
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \ \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

Proof. Omitted.

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

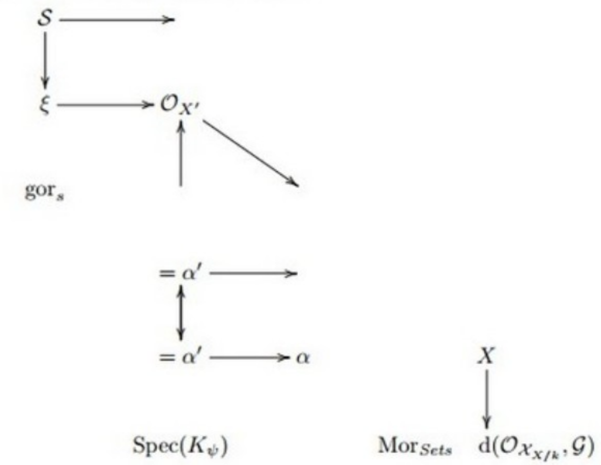
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

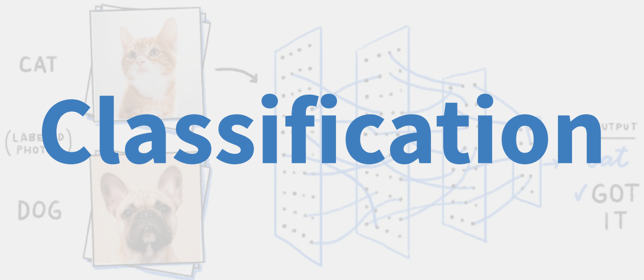
Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}^{-1}_{X_\ell}\mathcal{O}_{X_\lambda}(\mathcal{O}^{\overline{v}}_{X_\eta})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

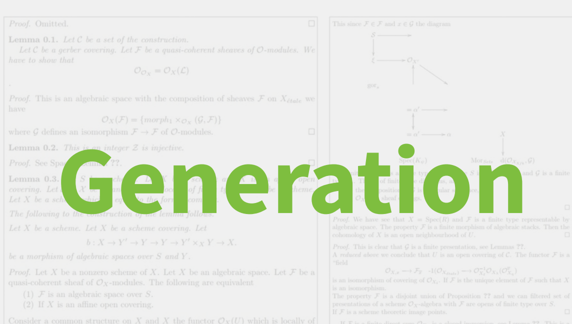
If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

Machine Learning at the LHC



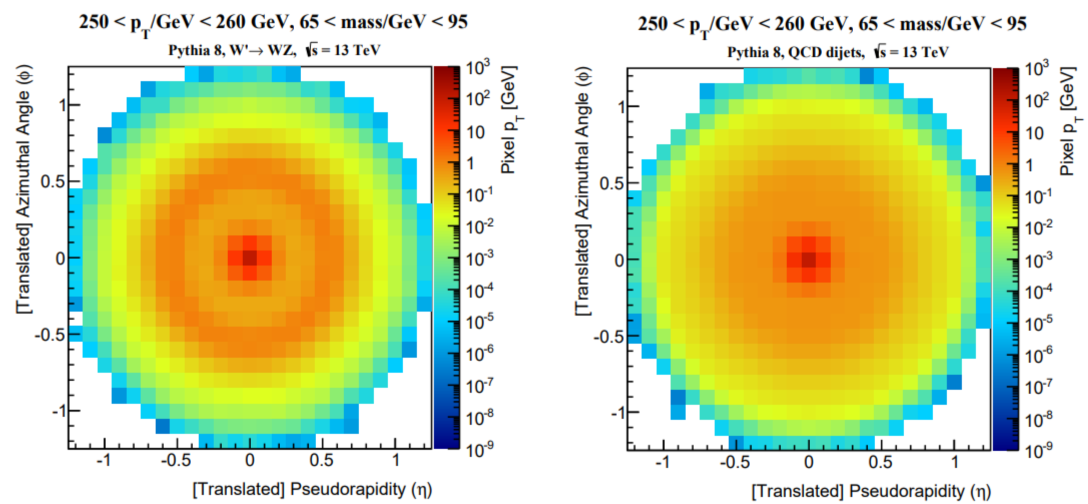
Classification

Regression



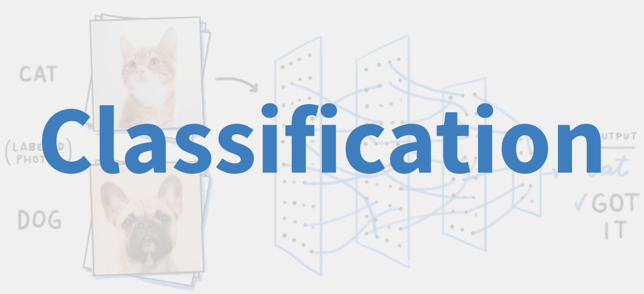
Generation

Classification: Jets



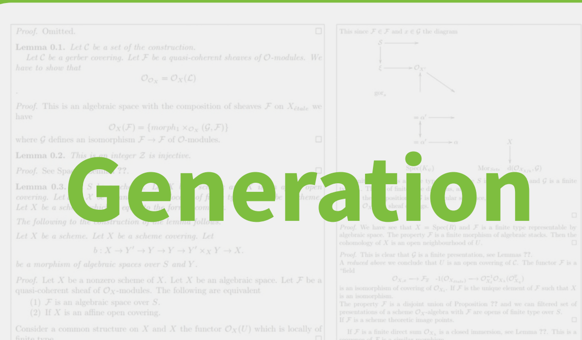
ArXiv: 1511.05190 L. Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman

Machine Learning at the LHC



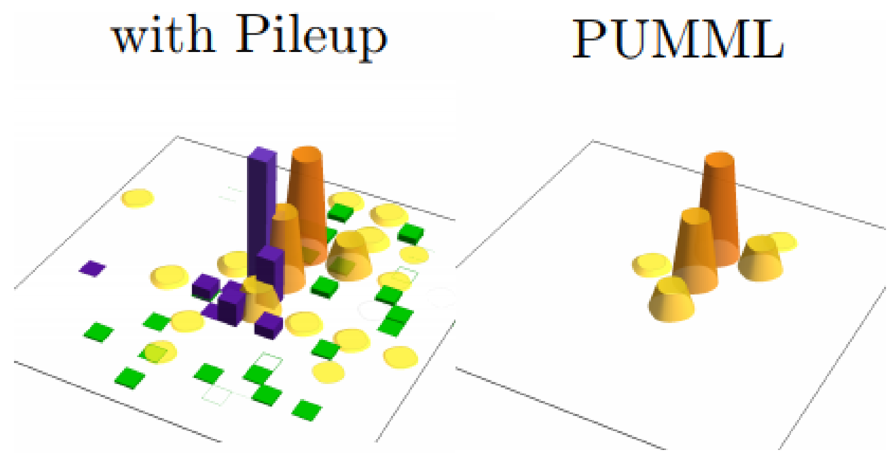
Classification

Regression



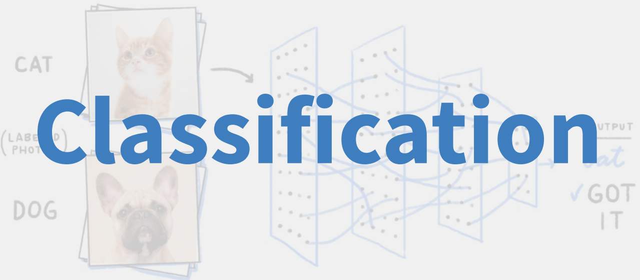
Generation

Regression: Pileup removal



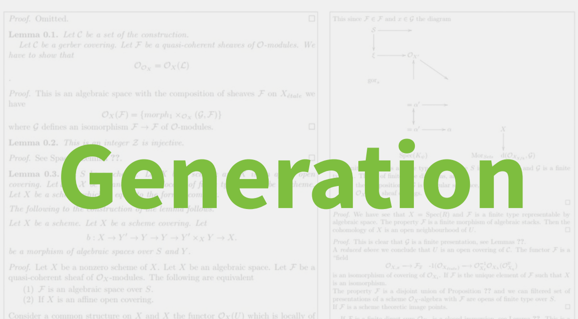
ArXiv:1707.08600. P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz

Machine Learning at the LHC



Classification

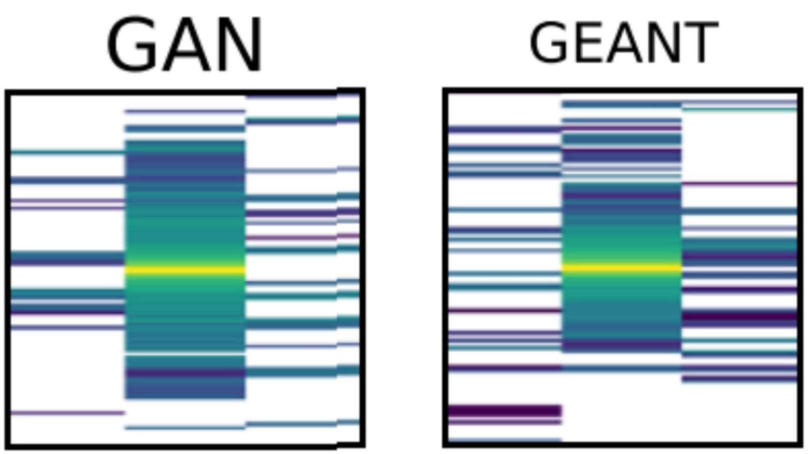
Regression



Generation

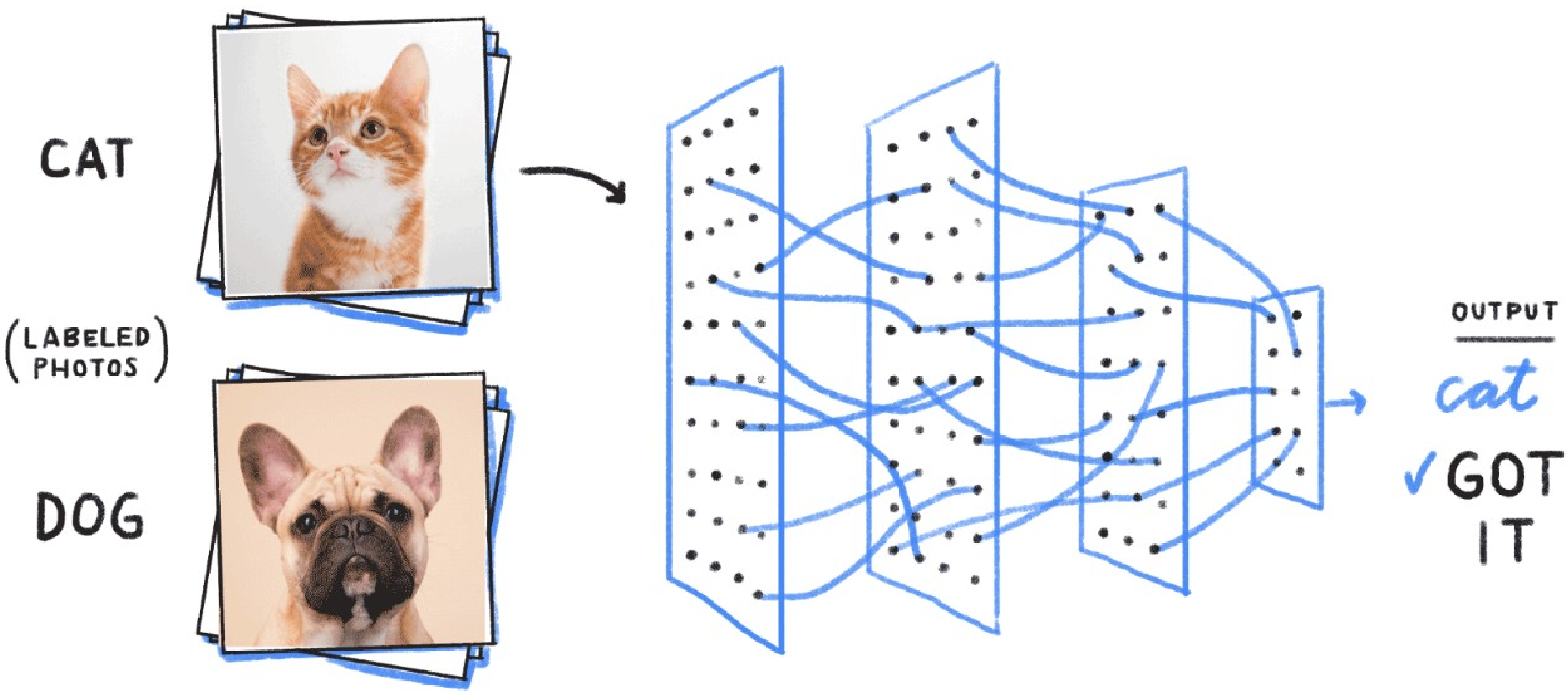
ArXiv 1712.10321, M. Paganini, L. de Oliveira, B. Nachamn

Generation: Fast simulation



Basic Machine Learning Primer

0) Decide objective: e.g. classify dog vs cat pictures
1) Choose a network architecture: e.g. CNN
2) Choose loss function (objective metric)
3) Train network using training data
4) Apply network on new test data



https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

Basic Machine Learning Primer

0) Decide objective: e.g. classify dog vs cat pictures

1) Choose a network architecture: e.g. CNN

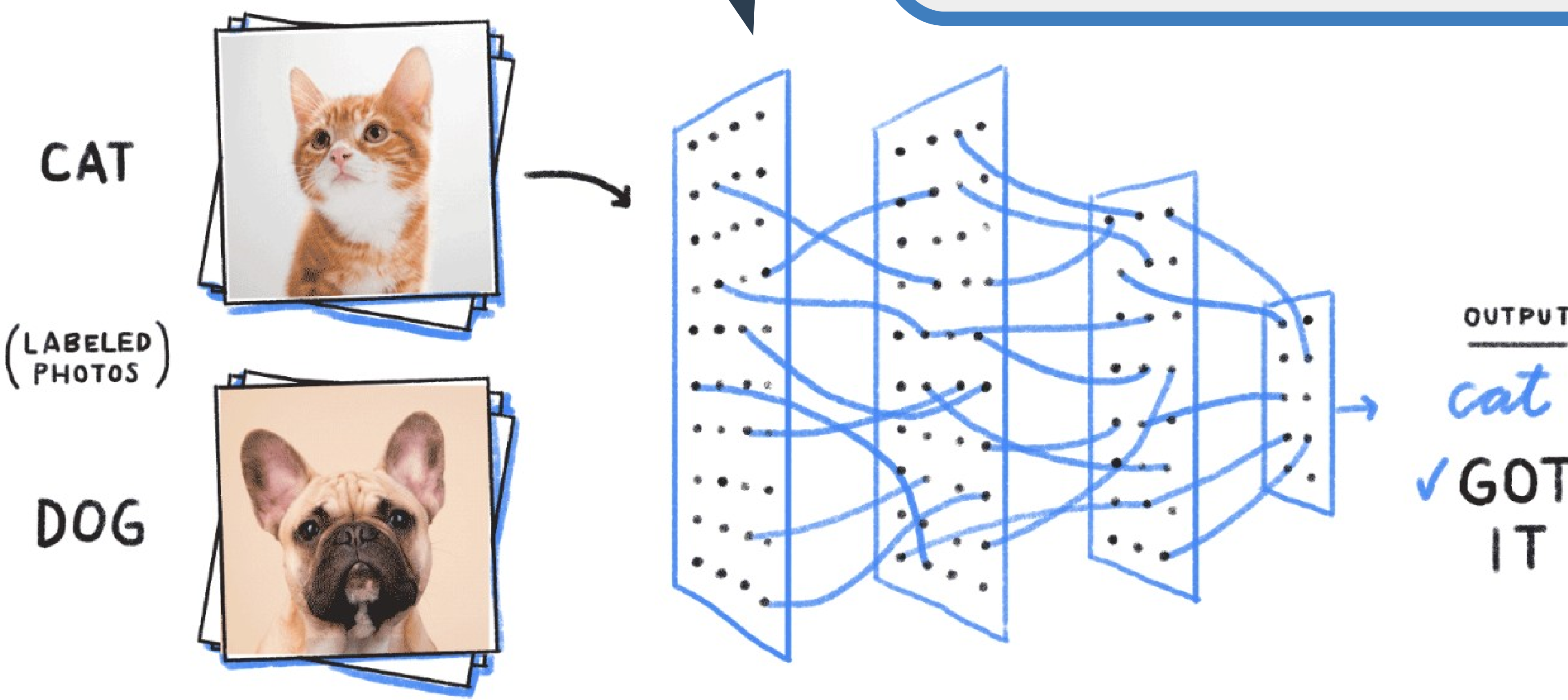
2) Choose loss function (objective metric)

3) Train network using training data

4) Apply network on new test data

A NN is just a function mapping N input numbers to M output numbers.

Trainable internal numbers – weights and biases – determine that function.



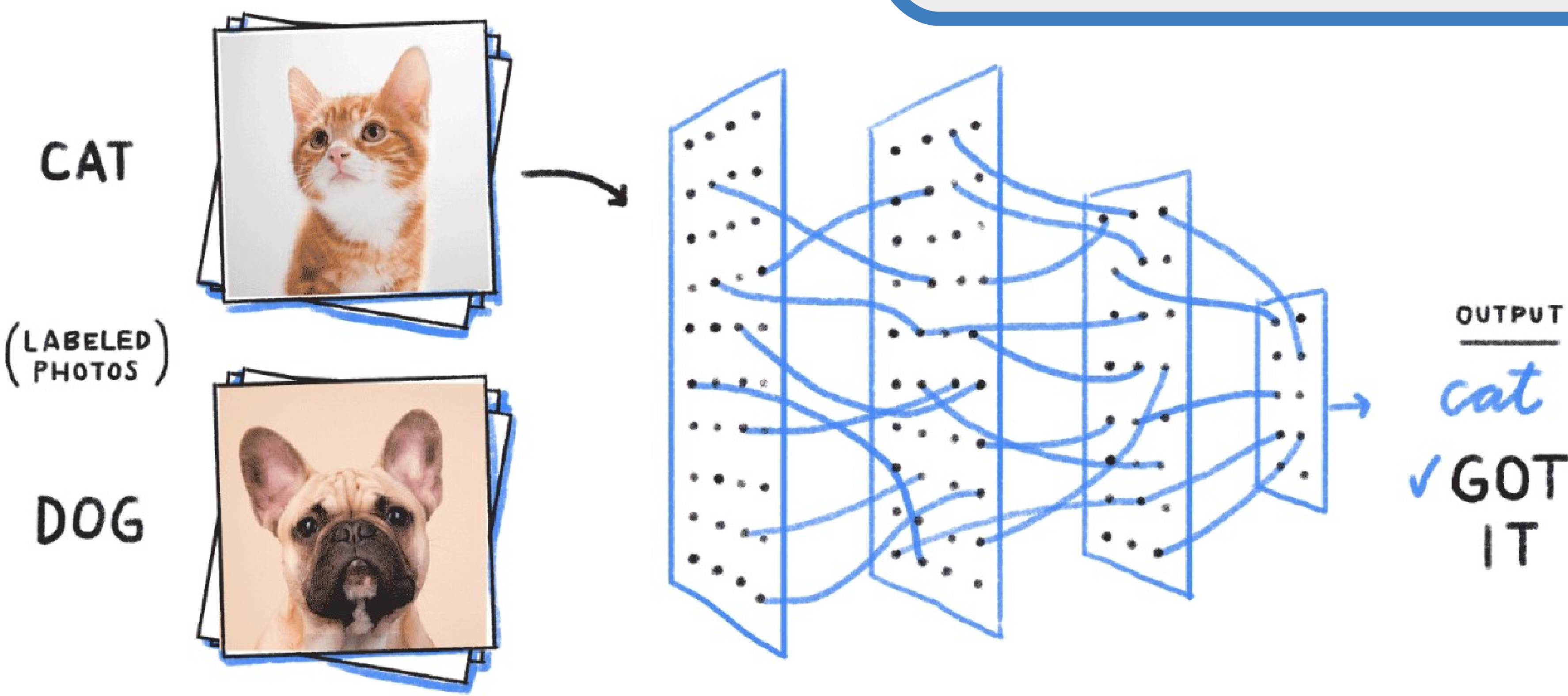
https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

Basic Machine Learning Primer

0) Decide objective: e.g. classify dog vs cat pictures
1) Choose a network architecture: e.g. CNN
2) Choose loss function (objective metric)
3) Train network using training data
4) Apply network on new test data

Naive example:
Fraction of correct predictions

Practical example: Cross-entropy loss
$-\sum y(x) \log[NN(x)]$



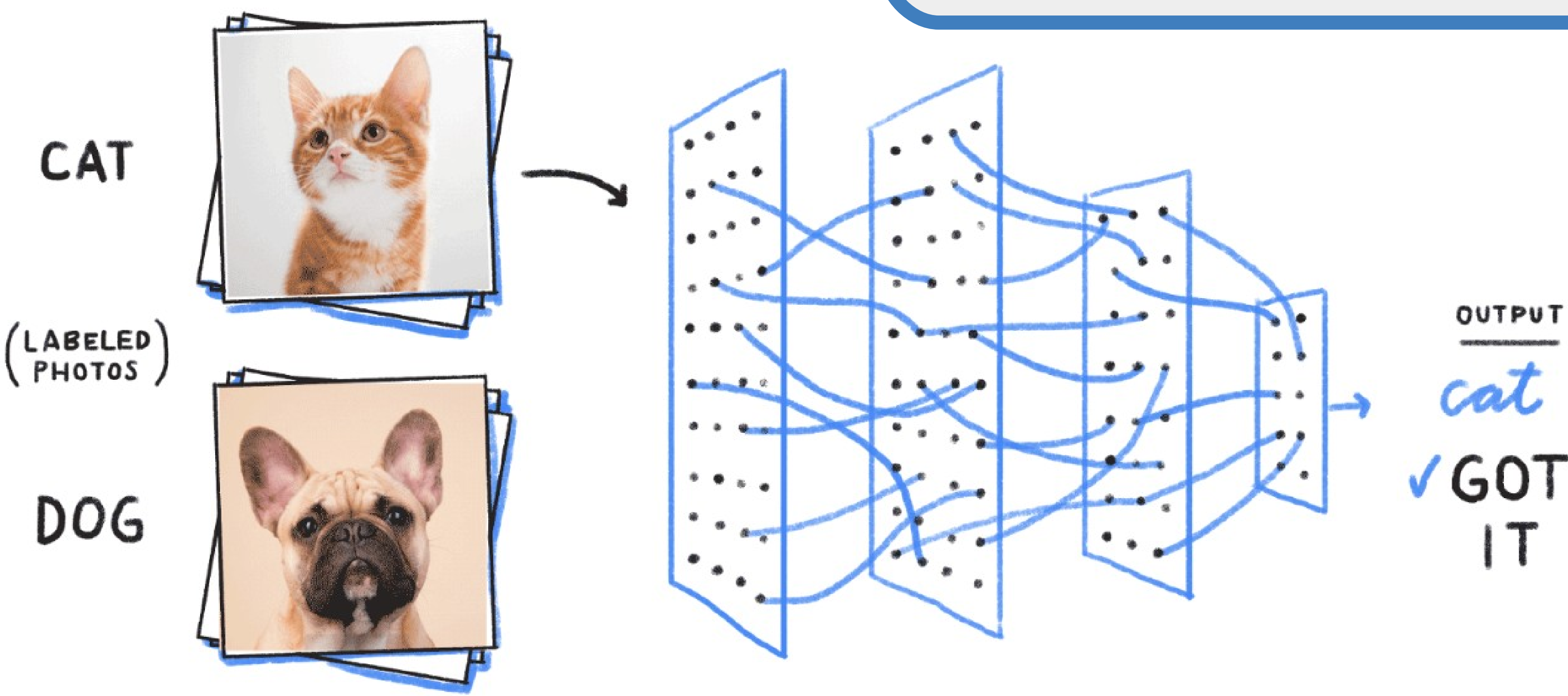
https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

Basic Machine Learning Primer

0) Decide objective: e.g. classify dog vs cat pictures
1) Choose a network architecture: e.g. CNN
2) Choose loss function (objective metric)
3) Train network using training data
4) Apply network on new test data

Use some iterative optimization algorithm to minimize the loss function on training data.

Be careful in selecting training data!



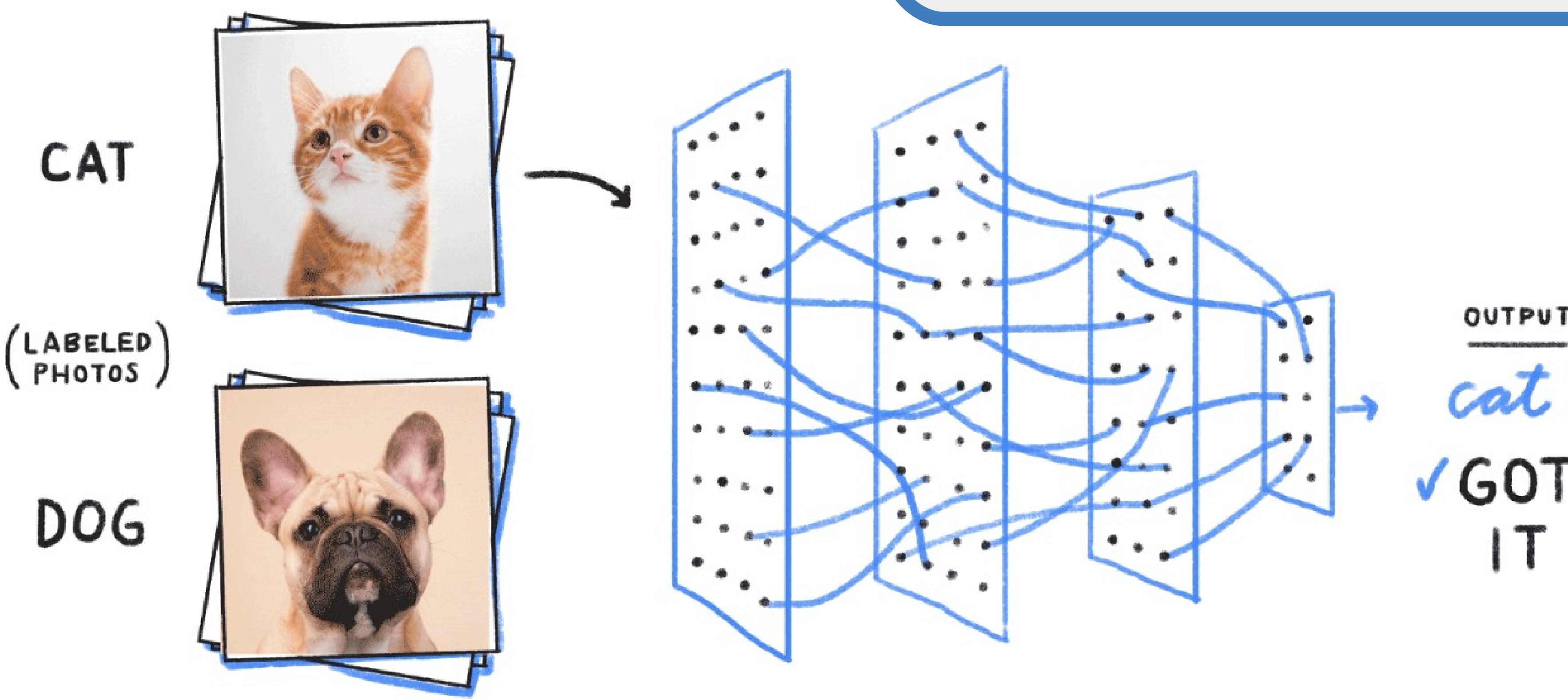
https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

Basic Machine Learning Primer

0) Decide objective: e.g. classify dog vs cat pictures
1) Choose a network architecture: e.g. CNN
2) Choose loss function (objective metric)
3) Train network using training data
4) Apply network on new test data

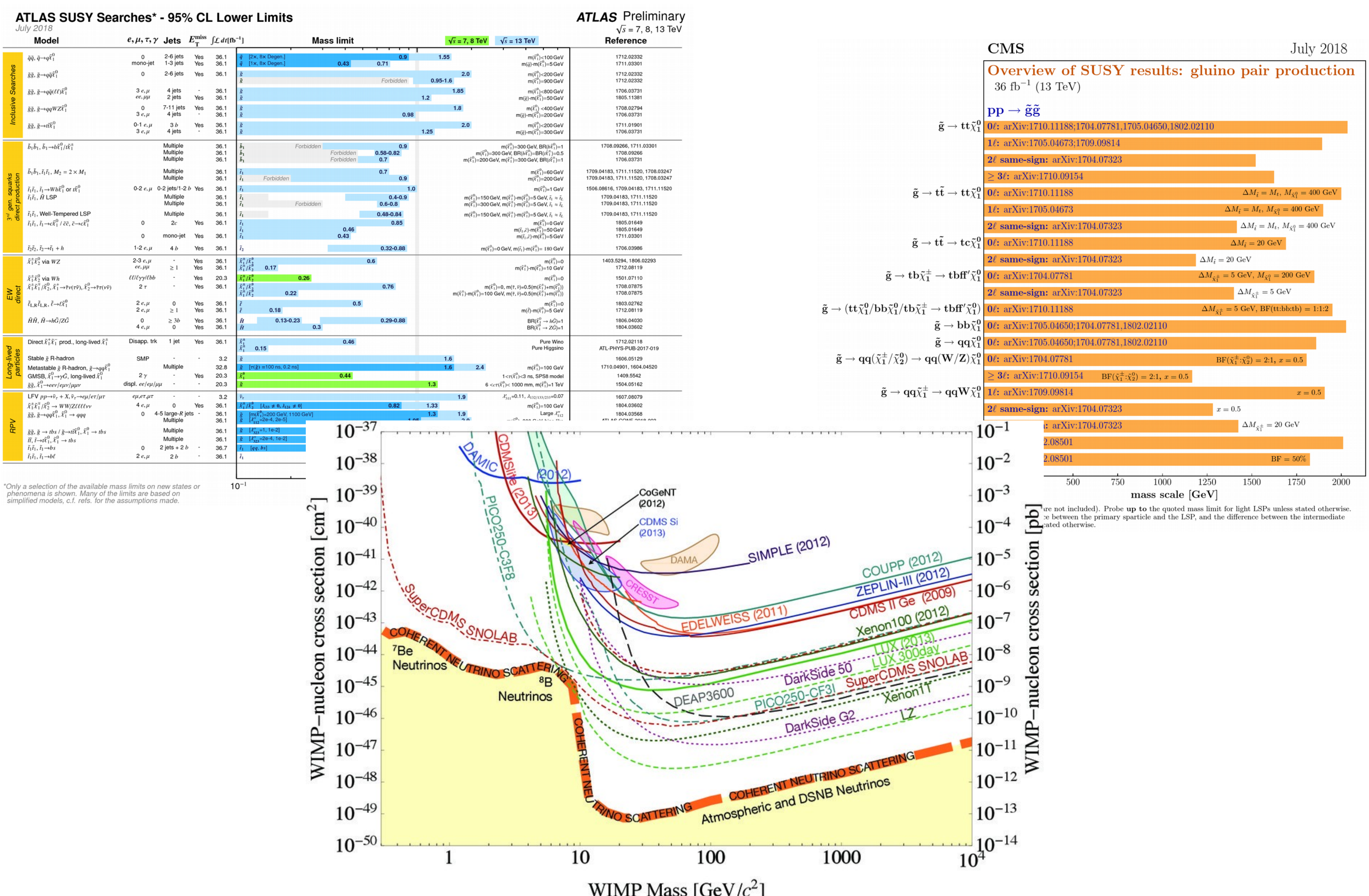
Training: **Slow**
Testing: **Fast**

Performance may be limited by the quality / relevance of training data.

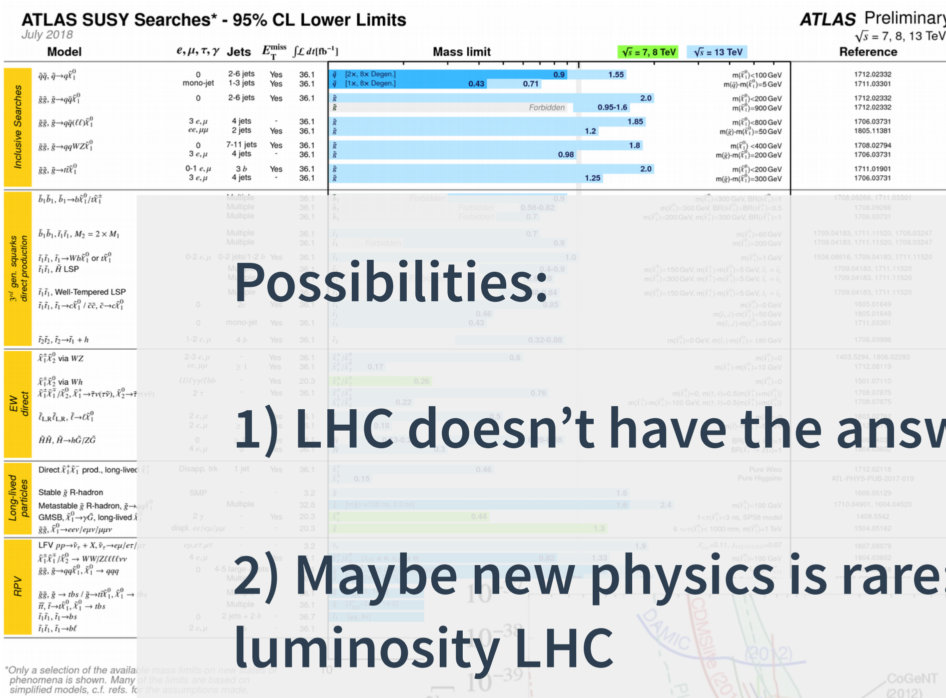


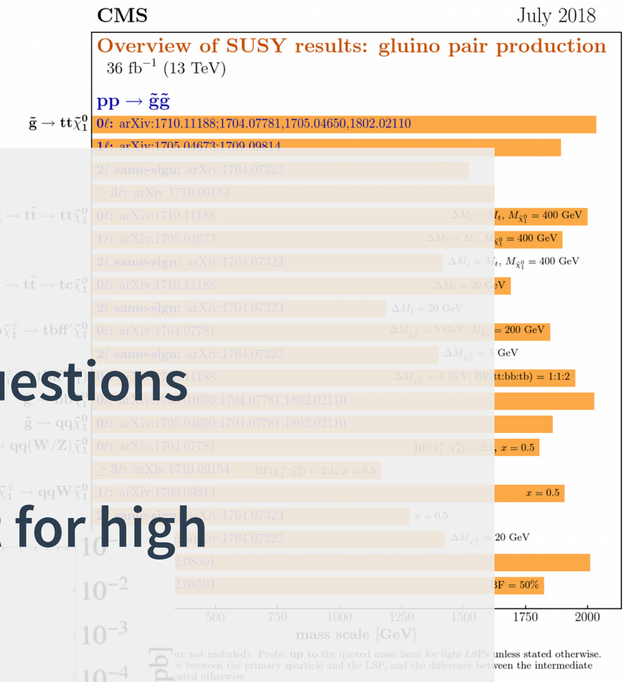
https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

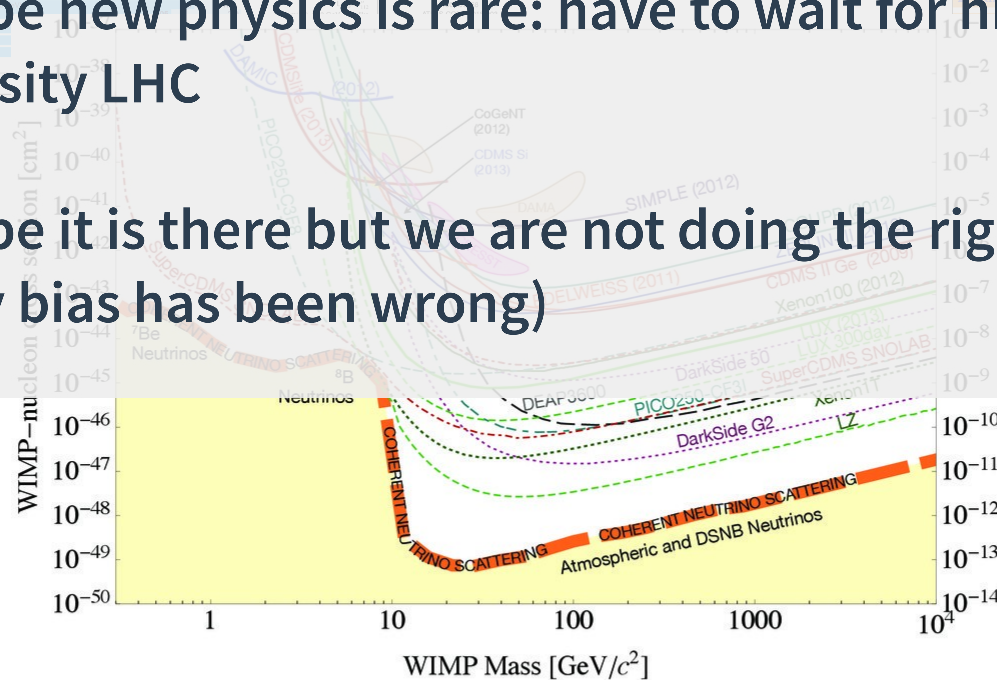
BSM Searches: Nothing so far



BSM Searches: Nothing so far







Possibilities:

1) LHC doesn't have the answers to our questions

2) Maybe new physics is rare: have to wait for high luminosity LHC

3) Maybe it is there but we are not doing the right search (theory bias has been wrong)

Are we missing something?

1) Ever-more sensitive dedicated searches for the standard culprits:

– Minimal Supersymmetry
– Top Partners
– diboson / ttbar resonances

Are we missing something?

1) Ever-more sensitive dedicated searches for the standard culprits:

- – Minimal Supersymmetry
- – Top Partners
- – diboson / ttbar resonances

2) General-purpose 'model-independent' searches for unexpected new physics

Signatures vs Models

E.g. 2-body resonances (pp → X → SM SM):

	e	μ	τ	γ	j	b	t	W	Z	h
e	$\pm\mp$[4],$\pm\pm$[5]	$\pm\pm$[5, 6] $\pm\mp$[6, 7]	[7]	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
μ		$\pm\mp$[4],$\pm\pm$[5]	[7]	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
τ			[8]	$\varnothing$	$\varnothing$	$\varnothing$	[9]	$\varnothing$	$\varnothing$	$\varnothing$
γ				[10]	[11–13]	$\varnothing$	$\varnothing$	[14]	[14]	$\varnothing$
j					[15]	[16]	[17]	[18]	[18]	$\varnothing$
b						[16]	[19]	$\varnothing$	$\varnothing$	$\varnothing$
t							[20]	[21]	$\varnothing$	$\varnothing$
W								[22–25]	[23, 24, 26, 27]	[28–30]
Z									[23, 25, 31]	[28, 30, 32, 33]
h										[34–37]

Signatures

[1610.09392] Craig, Draper, Kong, Ng, Whiteson

	e	μ	τ	γ	j	b	t	W	Z	h
e	$Z', H^{\pm\pm}$	$\not{R}, H^{\pm\pm}$	$\not{R}, H^{\pm\pm}$	L^*	$LQ, \not{R}$	$LQ, \not{R}$	$LQ, \not{R}$	L^*, ν_{KK}	L^*, e_{KK}	L^*
μ		$Z', H^{\pm\pm}$	$\not{R}, H^{\pm\pm}$	L^*	$LQ, \not{R}$	$LQ, \not{R}$	$LQ, \not{R}$	L^*, ν_{KK}	L^*, μ_{KK}	L^*
τ			$Z', H, H^{\pm\pm}$	L^*	$LQ, \not{R}$	$LQ, \not{R}$	$LQ, \not{R}$	L^*, ν_{KK}	L^*, τ_{KK}	L^*
γ				$H, G_{KK}, \mathcal{Q}$	Q^*	Q^*	Q^*	$W_{KK}, \mathcal{Q}$	$H, \mathcal{Q}$	Z_{KK}
j					Z', ρ, G_{KK}	$W', \not{R}$	$T', \not{R}$	Q^*, Q_{KK}	Q^*, Q_{KK}	Q'
b						Z', H	$W', \not{R}, H^{\pm}$	T', Q^*, Q_{KK}	Q^*, Q_{KK}	B'
t							H, G', Z'	T'	T'	T'
W								H, G_{KK}, ρ	$W', \mathcal{Q}$	$H^{\pm}, \mathcal{Q}, \rho$
Z									H, G_{KK}, ρ	A, ρ
h										H, G_{KK}

Models

Signatures vs Models

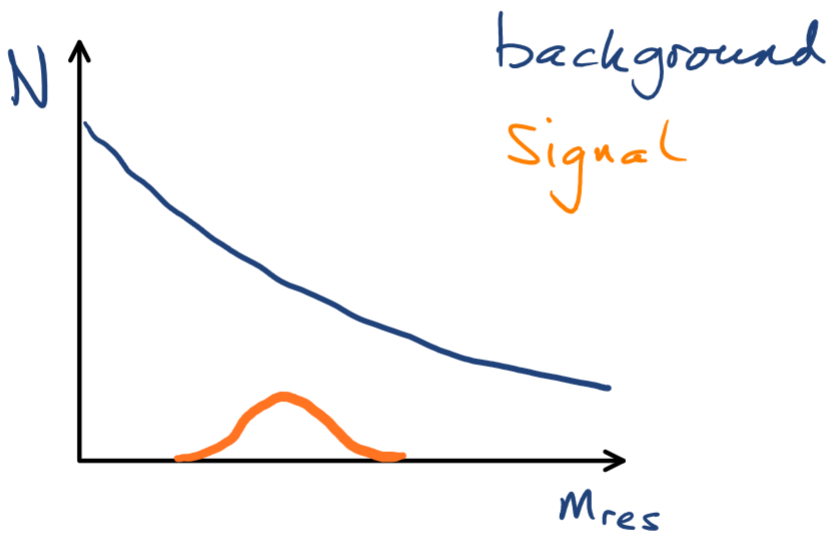
E.g. 2-body resonances:

	e	μ	τ	γ	j	b	t	W	Z	h	BSM
e	$\pm\mp$[4],$\pm\pm$[5]	$\pm\pm$[5, 6] $\pm\mp$[6, 7]	[7]	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	
μ		$\pm\mp$[4],$\pm\pm$[5]	[7]	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	
τ			[8]	$\varnothing$	$\varnothing$	$\varnothing$	[9]	$\varnothing$	$\varnothing$	$\varnothing$	
γ				[10]	[11–13]	$\varnothing$	$\varnothing$	[14]	[14]	$\varnothing$	
j					[15]	[16]	[17]	[18]	[18]	$\varnothing$	
b						[16]	[19]	$\varnothing$	$\varnothing$	$\varnothing$	
t							[20]	[21]	$\varnothing$	$\varnothing$	
W								[22–25]	[23, 24, 26, 27]	[28–30]	
Z									[23, 25, 31]	[28, 30, 32, 33]	
h										[34–37]	
BSM											

pp → X → BSM BSM largely uncovered

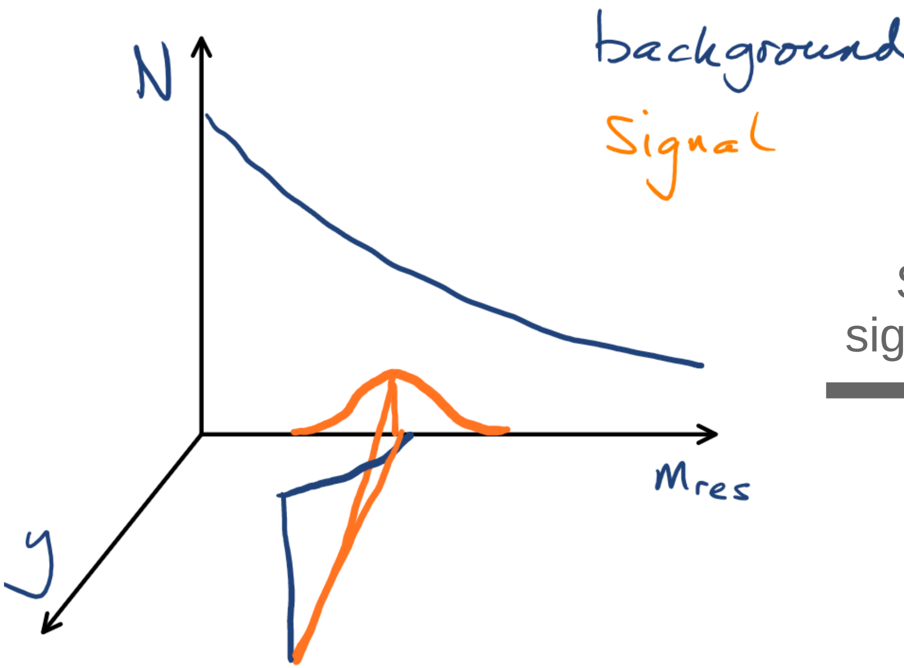
Basic Resonance Searches

E.g. Dijet Search



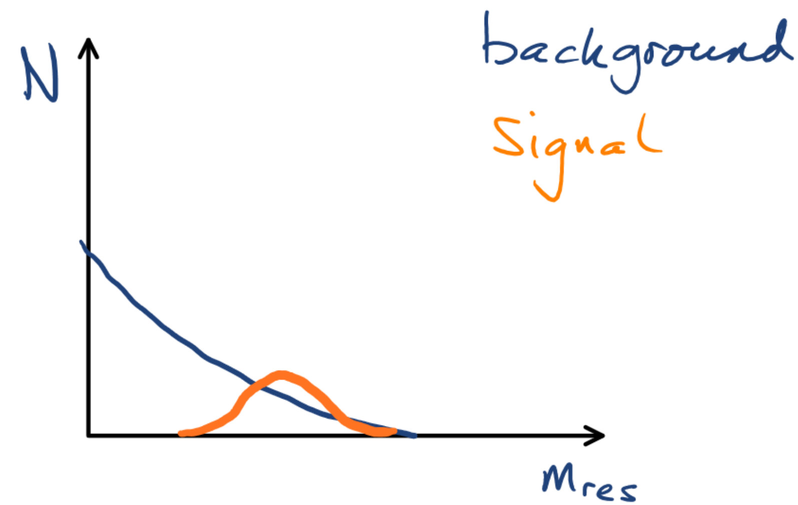
Basic Resonance Searches

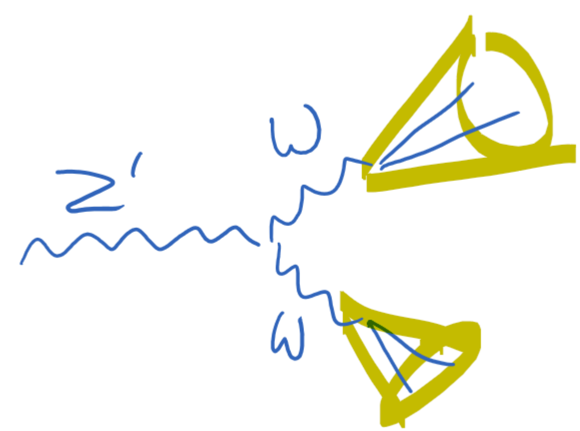
E.g. Dijet Search



Selection for signal-like events

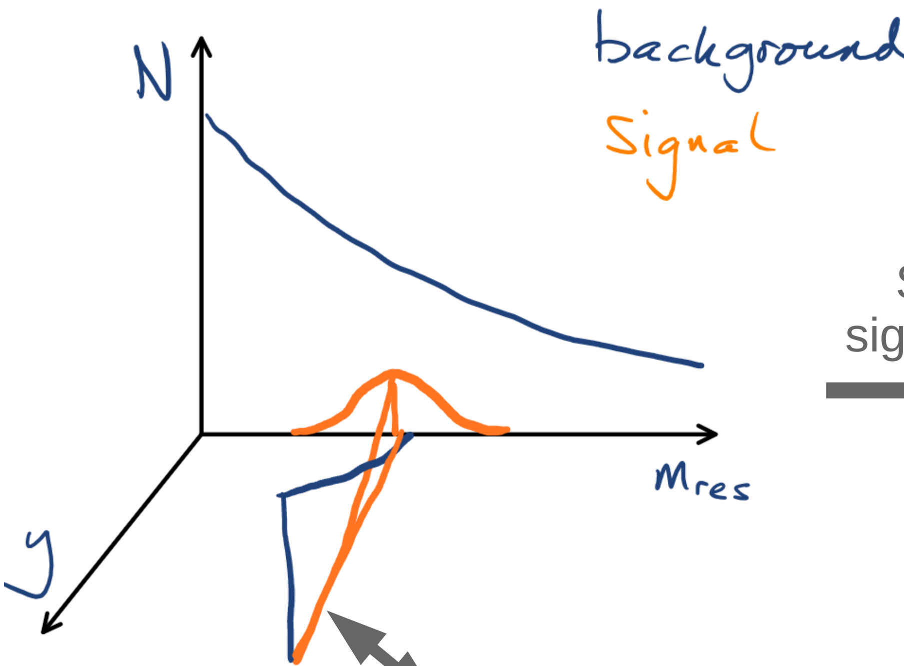
E.g. WW resonance, tt resonance, etc.





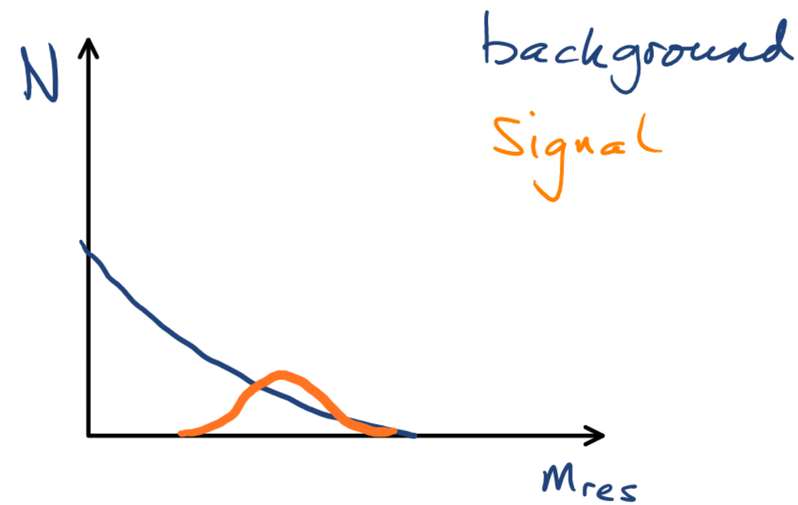
Basic Resonance Searches

E.g. Dijet Search

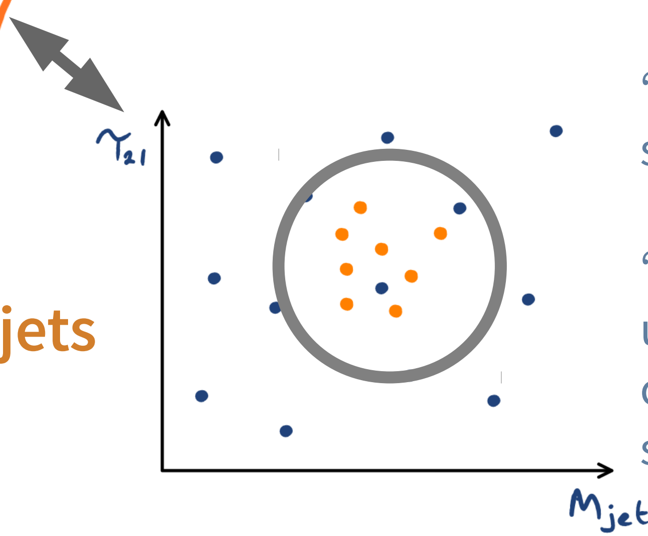


E.g. WW resonance, tt resonance, etc.

Selection for signal-like events



W-jets

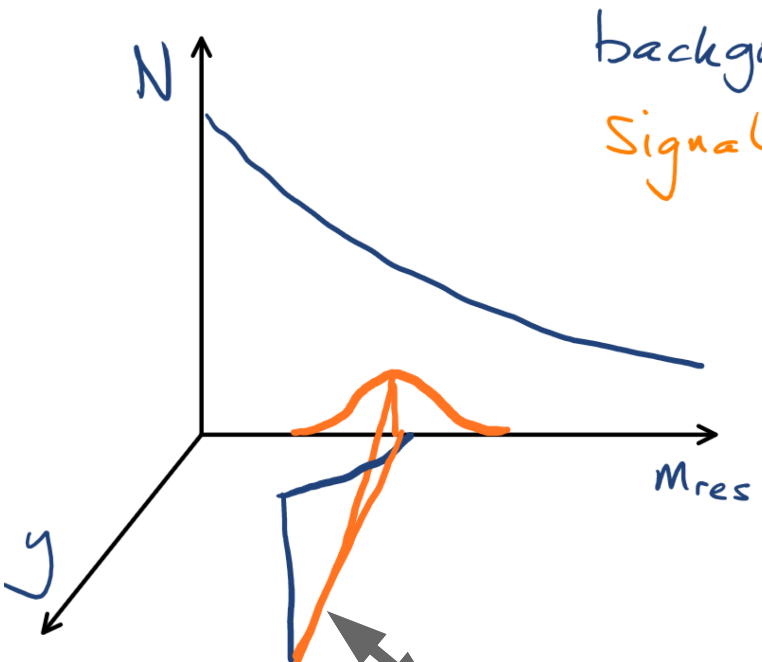


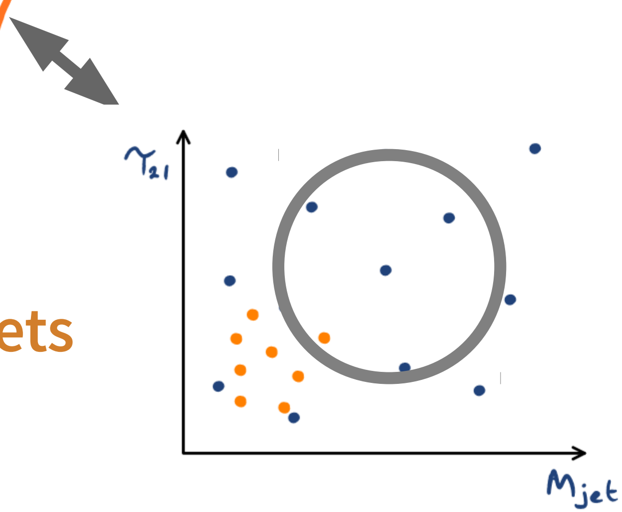
'Old fashioned': Simple few-D substructure selection

'Modern': Deep NN classifier using ~few hundred jet constituent inputs (~300-D selection).

Basic Resonance Searches

E.g. Dijet Search

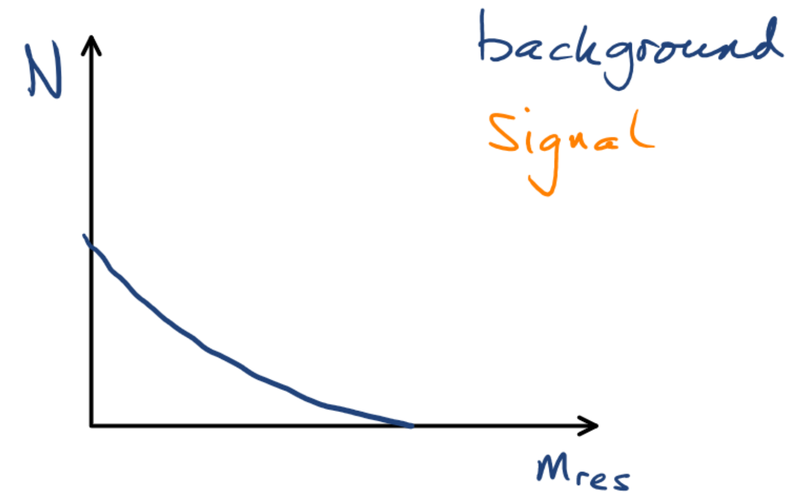


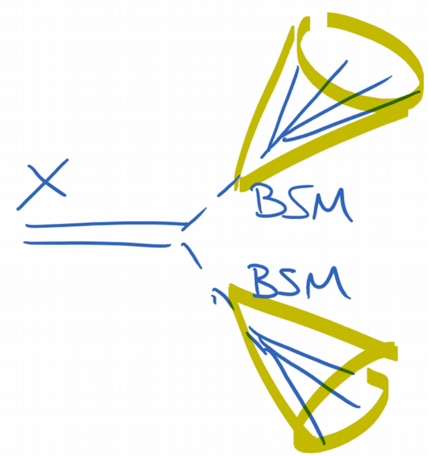


?-jets

Selection for
signal-like events

E.g. ?? resonance





A Traditional Dichotomy

Model Inclusive Search

– Weak signal assumptions
– Basic selection criteria in few variables
– Large backgrounds

– Risk missing a signal under background

Model Specific Search

– Strong signal assumptions
– Sophisticated multivariate selection
– Small backgrounds

– Risk not making the 'correct' signal selection

How to make a search with a sophisticated multivariate selection to beat backgrounds while using weak signal assumptions (unknown specific signal model)?

→ Learn selection from data

Why Train Machines on Data?

1) Maybe you have not simulated the correct signal model **(either because you haven't thought of it, or because it involves non-perturbative physics that prevents simulation)**

Why Train Machines on Data?

1) Maybe you have not simulated the correct signal model (either because you haven't thought of it, or because it involves non-perturbative physics that prevents simulation)

2) Monte-Carlo simulation for training data may differ from real LHC data

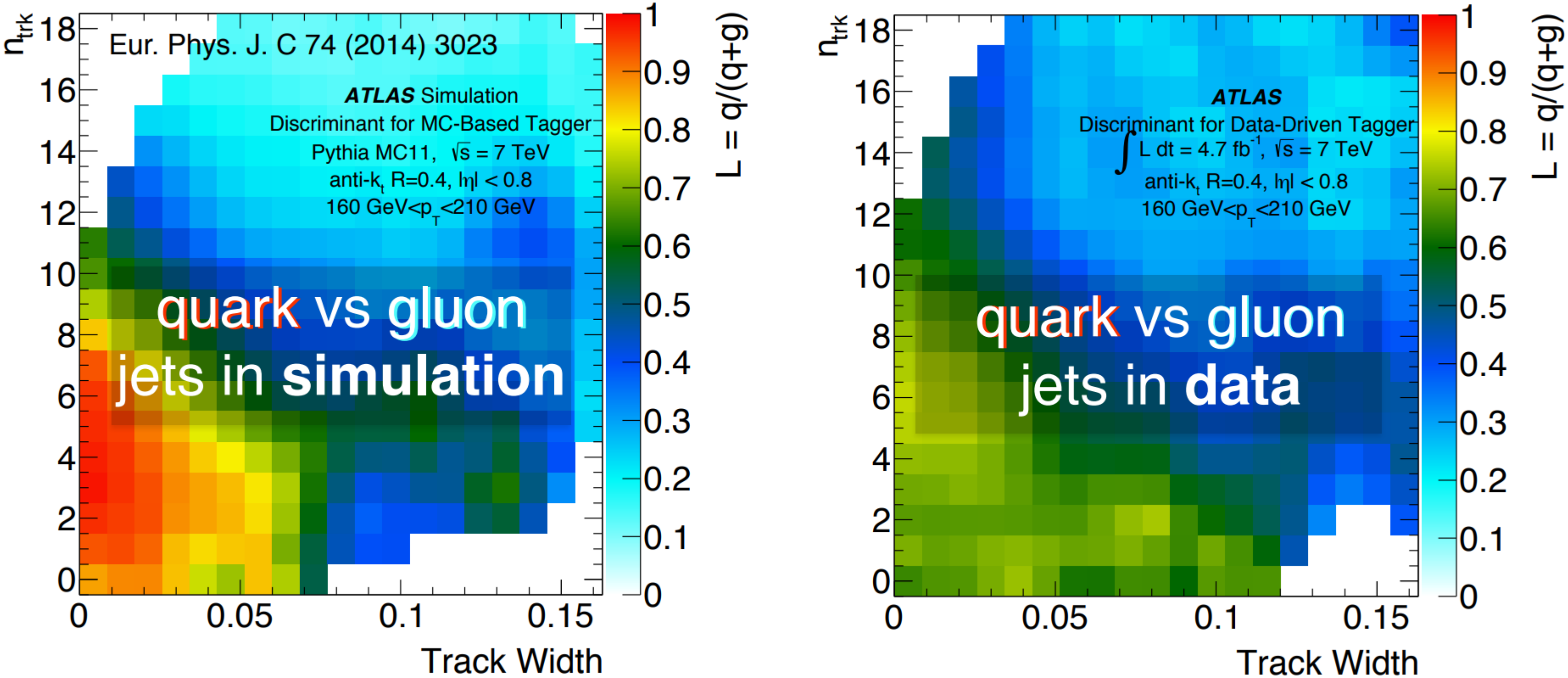
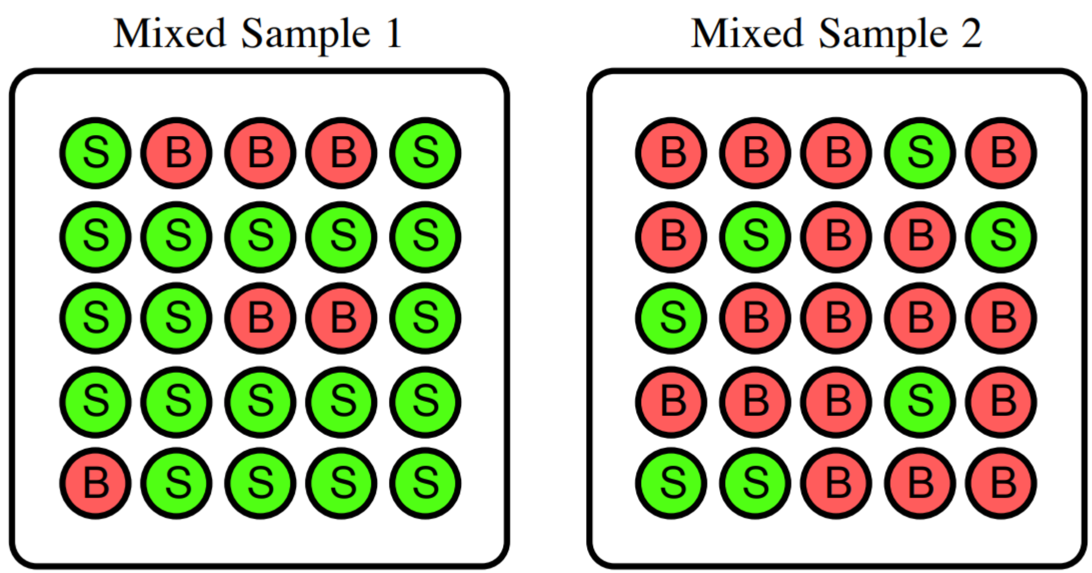
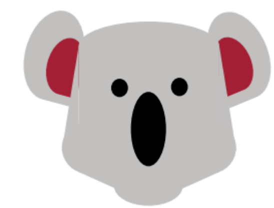


Figure taken from Ben Nachman's talk at BOOST 2018
https://indico.cern.ch/event/649482/contributions/2993322/attachments/1688082/2715256/WeakSupervision_BOOST2018.pdf

Solution for ML:

Train directly on data using mixed samples





A) LoLiProp (Learning from Labelled Proportions)
Train using class proportions

[1702.00414] L. Dery, B. Nachman, F. Rubbo, A. Schwartzman
[1706.09451] T. Cohen, M. Freytsis, B. Ostdiek

B) CWoLa (Classification Without Labels)

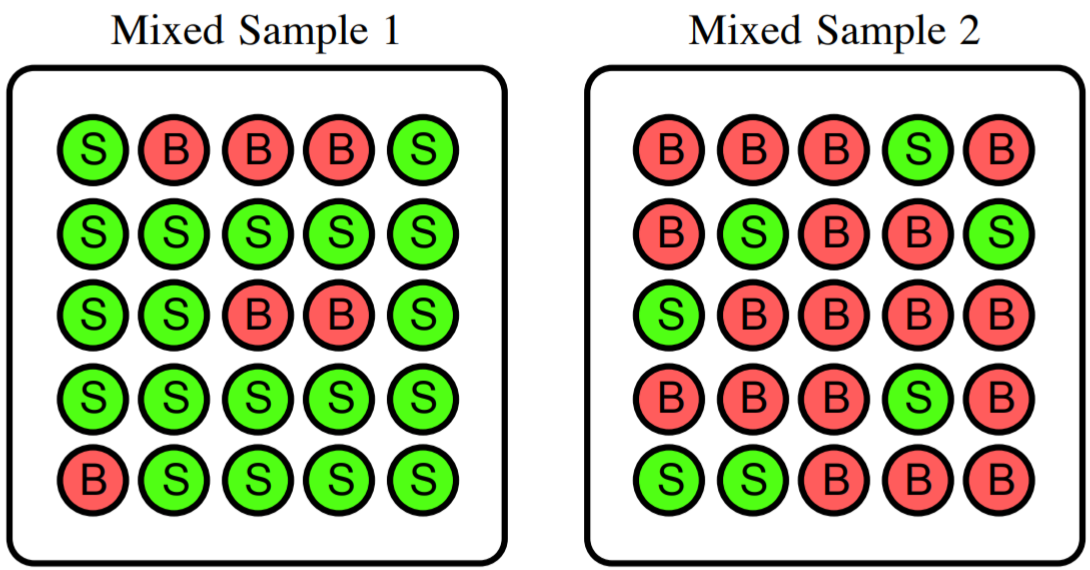
Train to classify as mixed sample 1 or 2.

[1708.02949] E. M. Metodiev, B. Nachman, J. Thaler

P. T. Komiske, E. M. Metodiev, B. Nachman, M. D. Schwartz
See also [1801.10158]

CWoLa

CWoLa

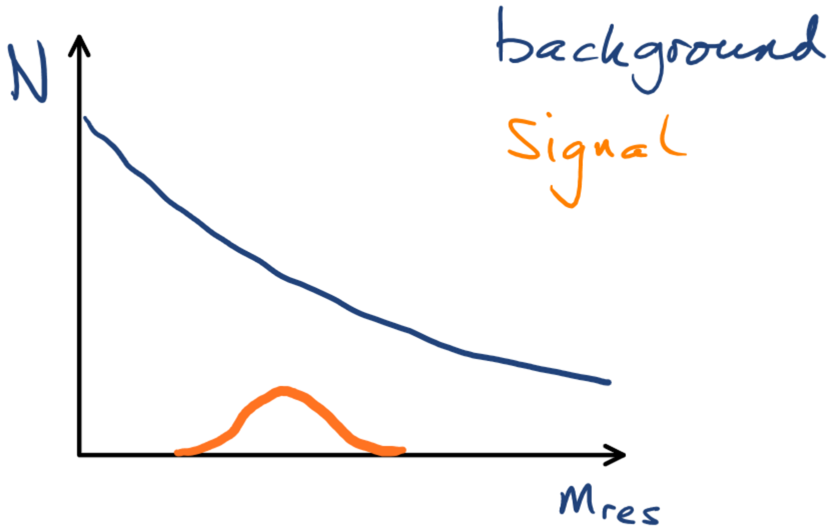


Classifier trained to optimally discriminate mixed sample 1 from mixed sample 2 *is also optimal* for discriminating S from B, so long as:

– Samples 1 and 2 contain different fractions of S and B
– S in sample 1 is drawn from the same distribution as S in sample 2
– B in sample 1 is drawn from the same distribution as B in sample 2
– Training statistics are sufficiently large

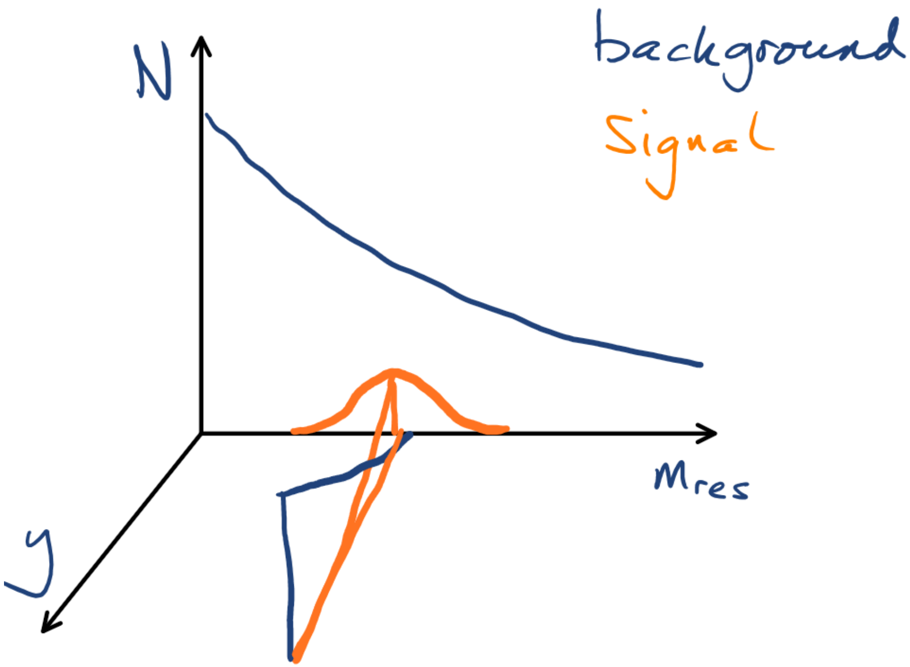
How to use this for a search where S is new physics and B is SM background?

CWoLa Hunting



1. Assume signal is localized in some specific variable in which background is smooth.

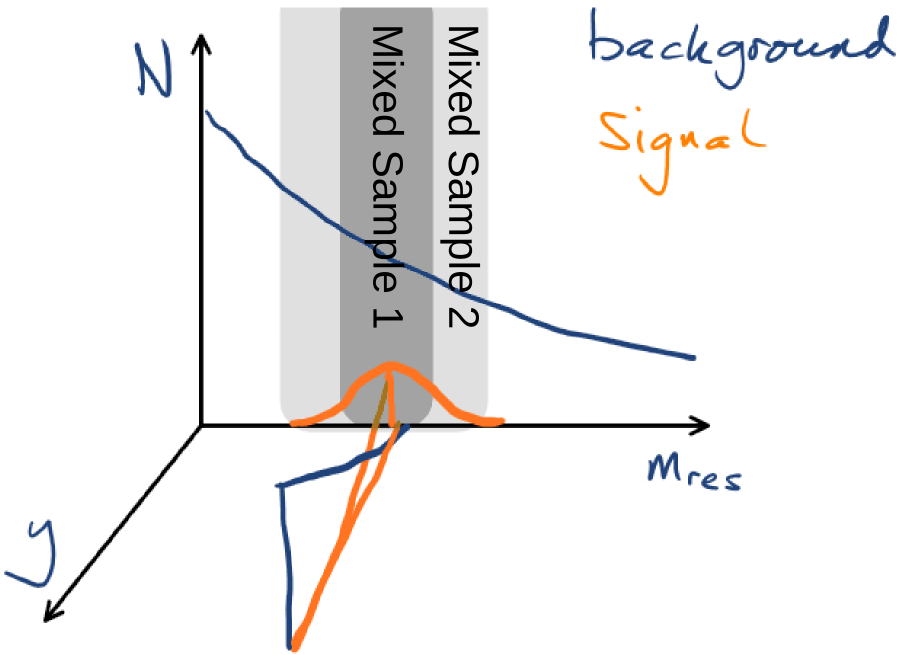
CWoLa Hunting



1. Assume signal is localized in some specific variable in which background is smooth.

2. Assume signal has some distinguishing characteristics within some broad set of additional observables *y*.

CWoLa Hunting

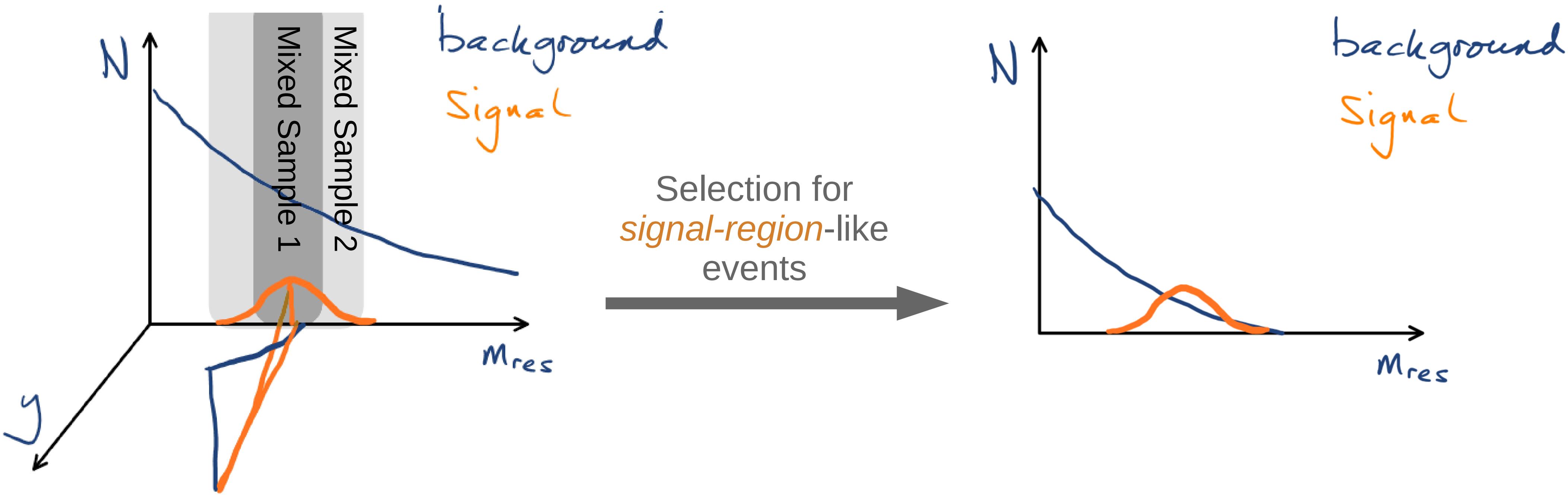


1. Assume signal is localized in some specific variable in which background is smooth.

2. Assume signal has some distinguishing characteristics within some broad set of additional observables *y*.

3. For some resonance mass hypothesis, split data into signal-region and sideband-region mixed samples

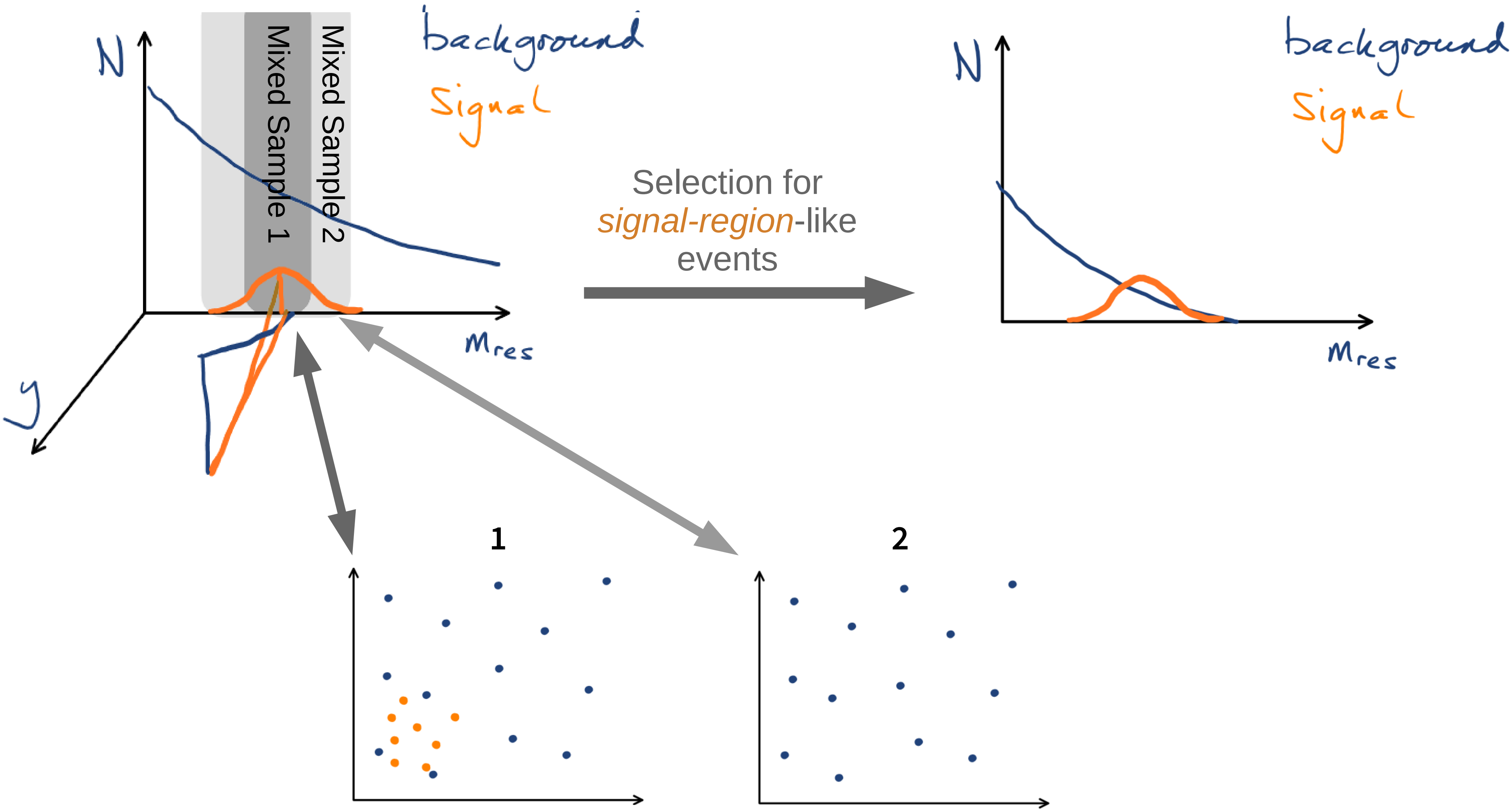
CWoLa Hunting



Train classifier to discriminate samples based on variables *y*

Note: background *y* distribution should not be strongly varying with the resonance variable.

CWoLa Hunting

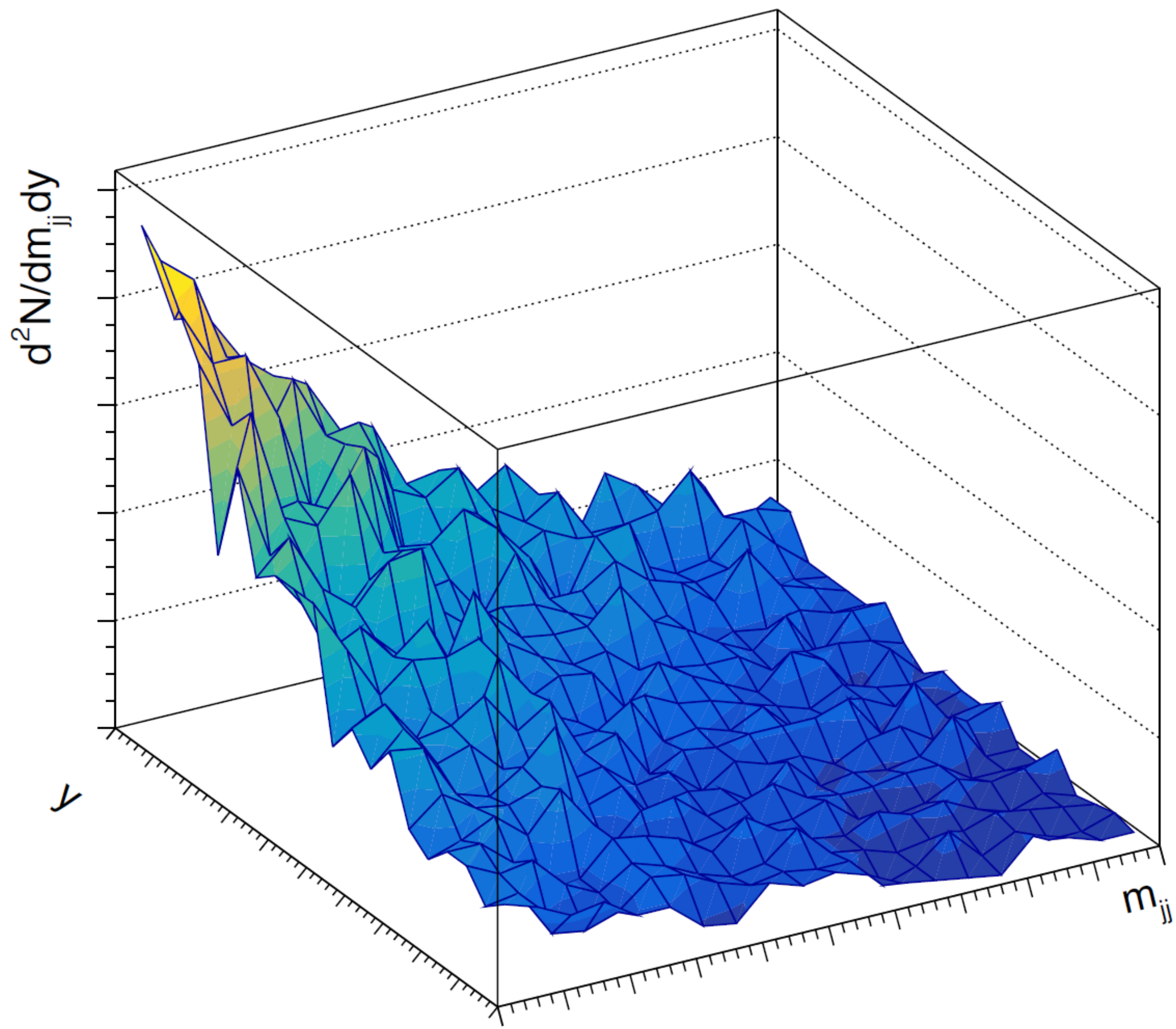


Overfitting and the Look Elsewhere Effect

Of course, there is going to be a large trials factor, especially if *y* is high-dimensional.

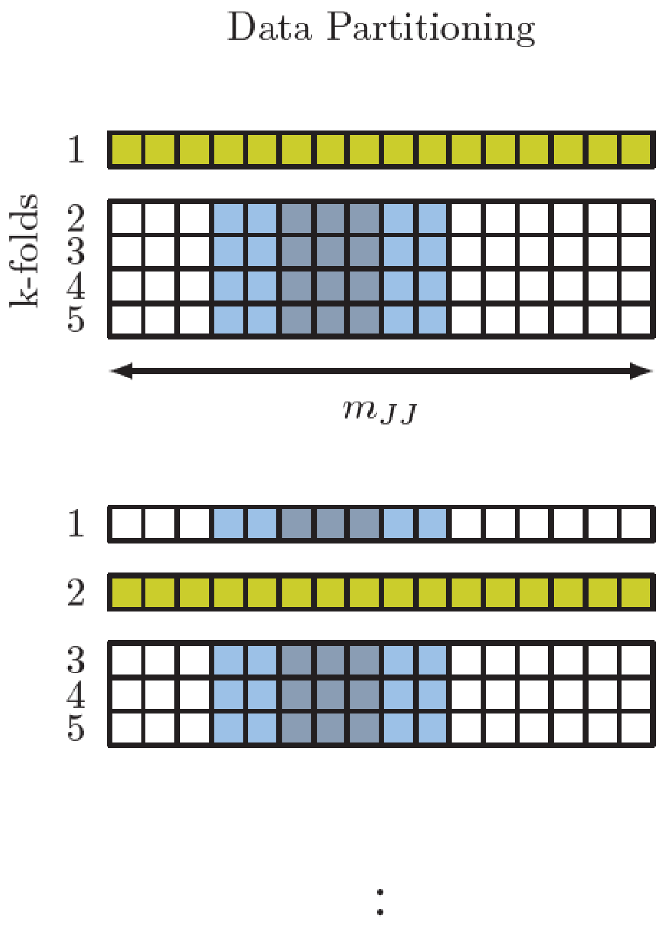
Easy solution:
Train test split
(Statistical fluctuations in training and test set are uncorrelated)

More sophisticated:
Nested cross-training



Nested Cross-Training

1) Divide entire dataset into k-folds



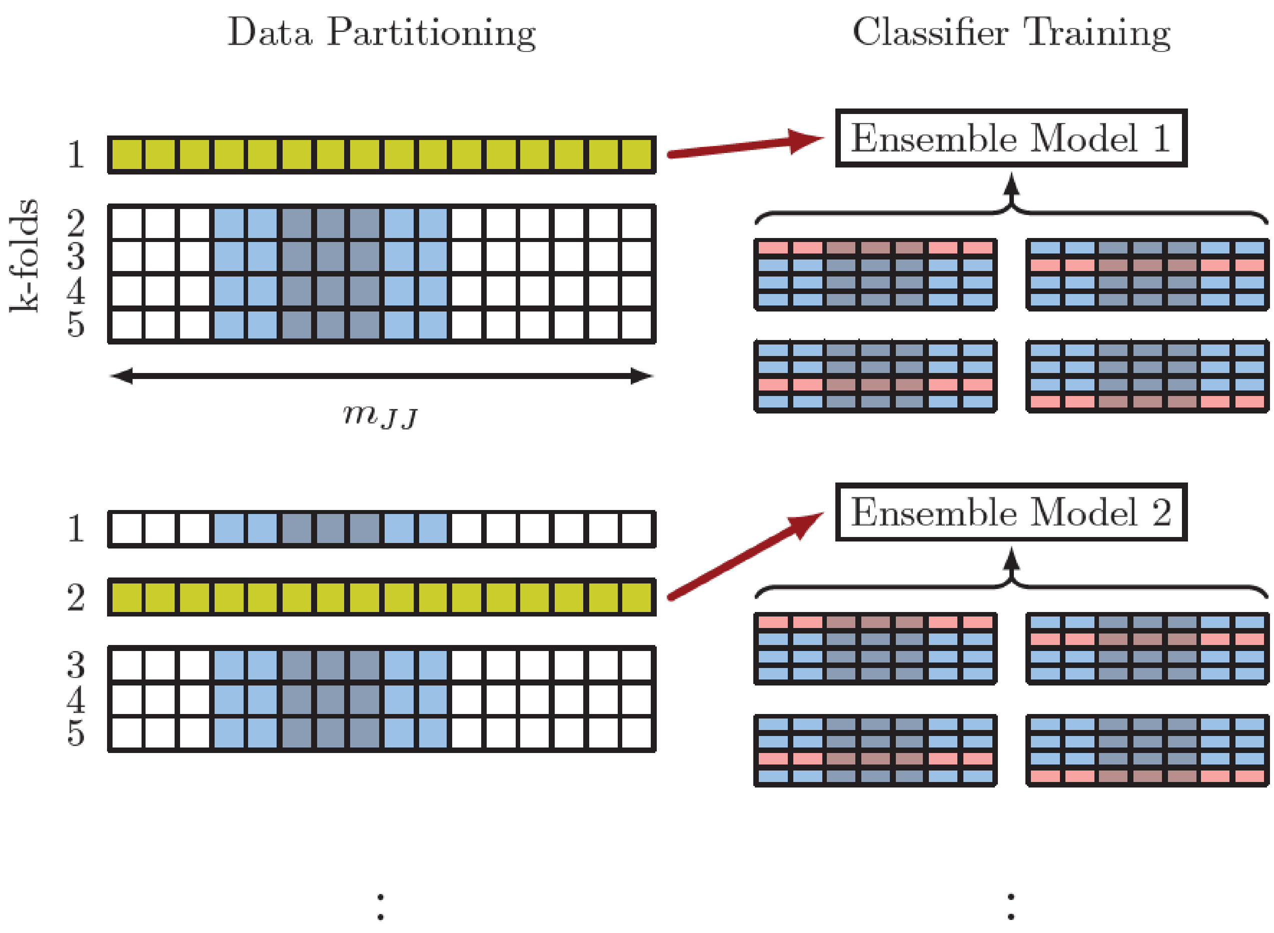
Test set

Training signal region

Training sideband

Nested Cross-Training

2) Train CWoLa Classifiers



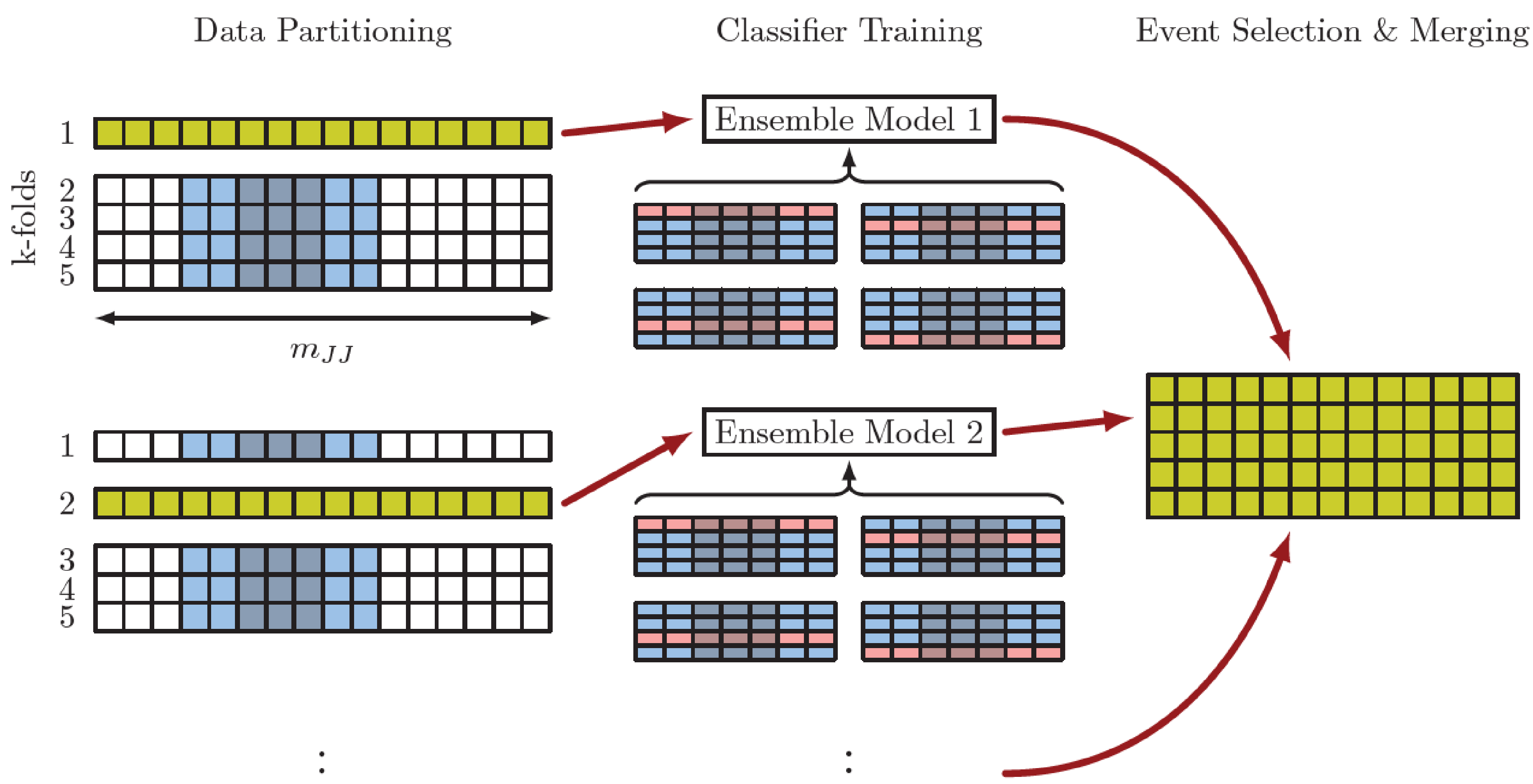
Train **signal** vs **sideband** k-1 times, rotating **validation set**.

Average the k-1 models to form an ensemble model

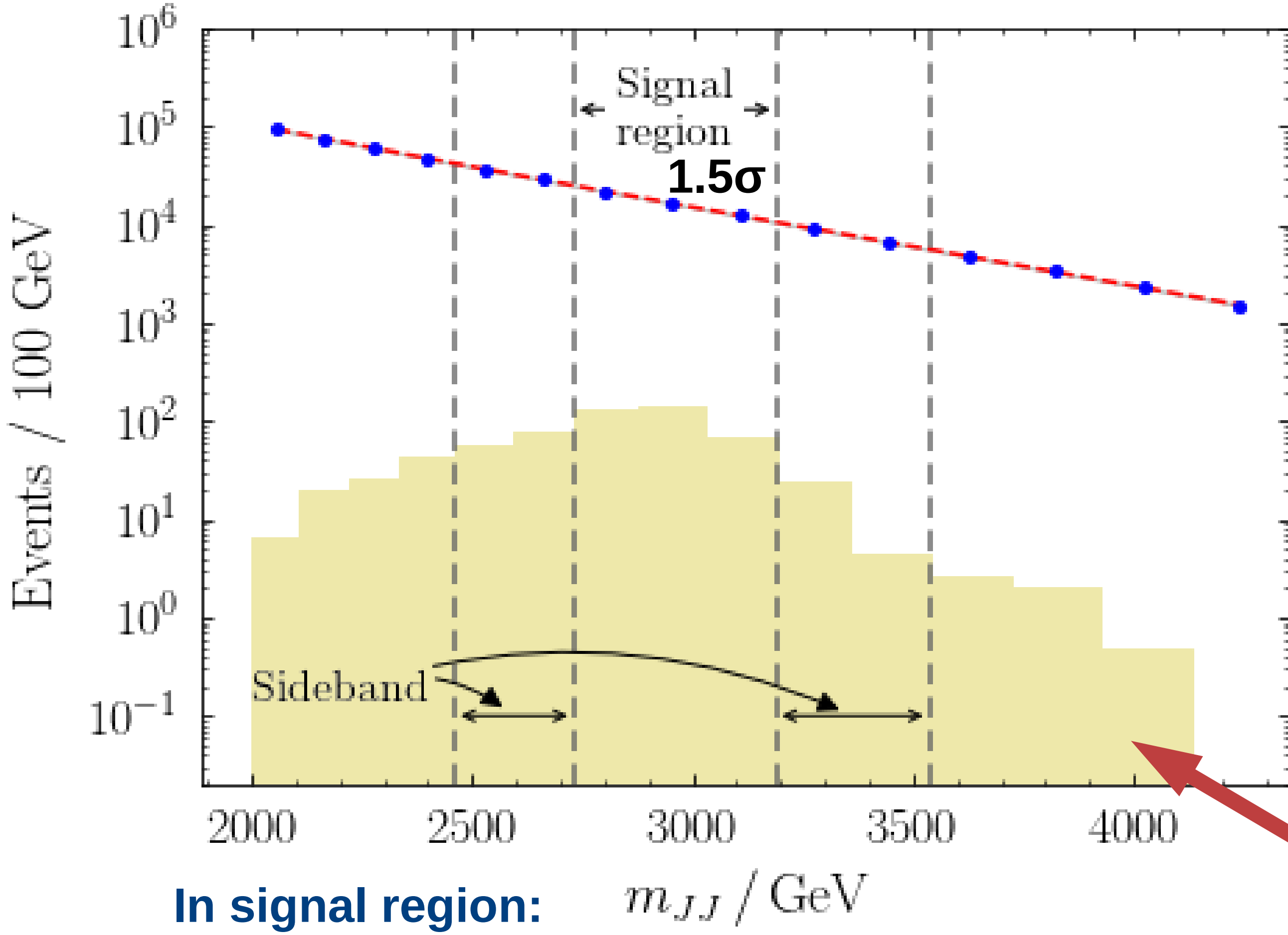
Background fluctuation contributions will destructively interfere, signal contributions constructively.

Nested Cross-Training

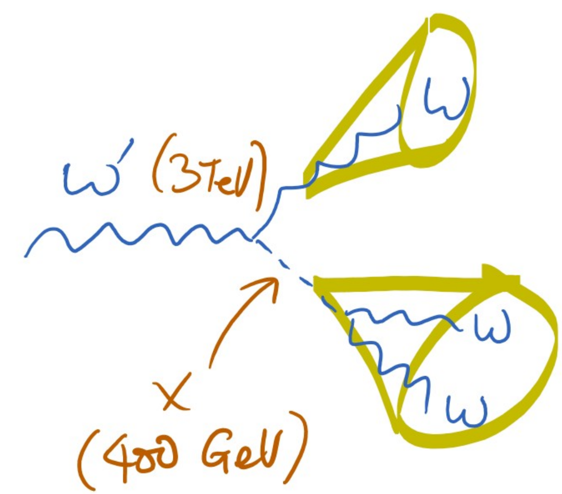
3) Select events in each k-fold and then merge



Application to Bump Hunt

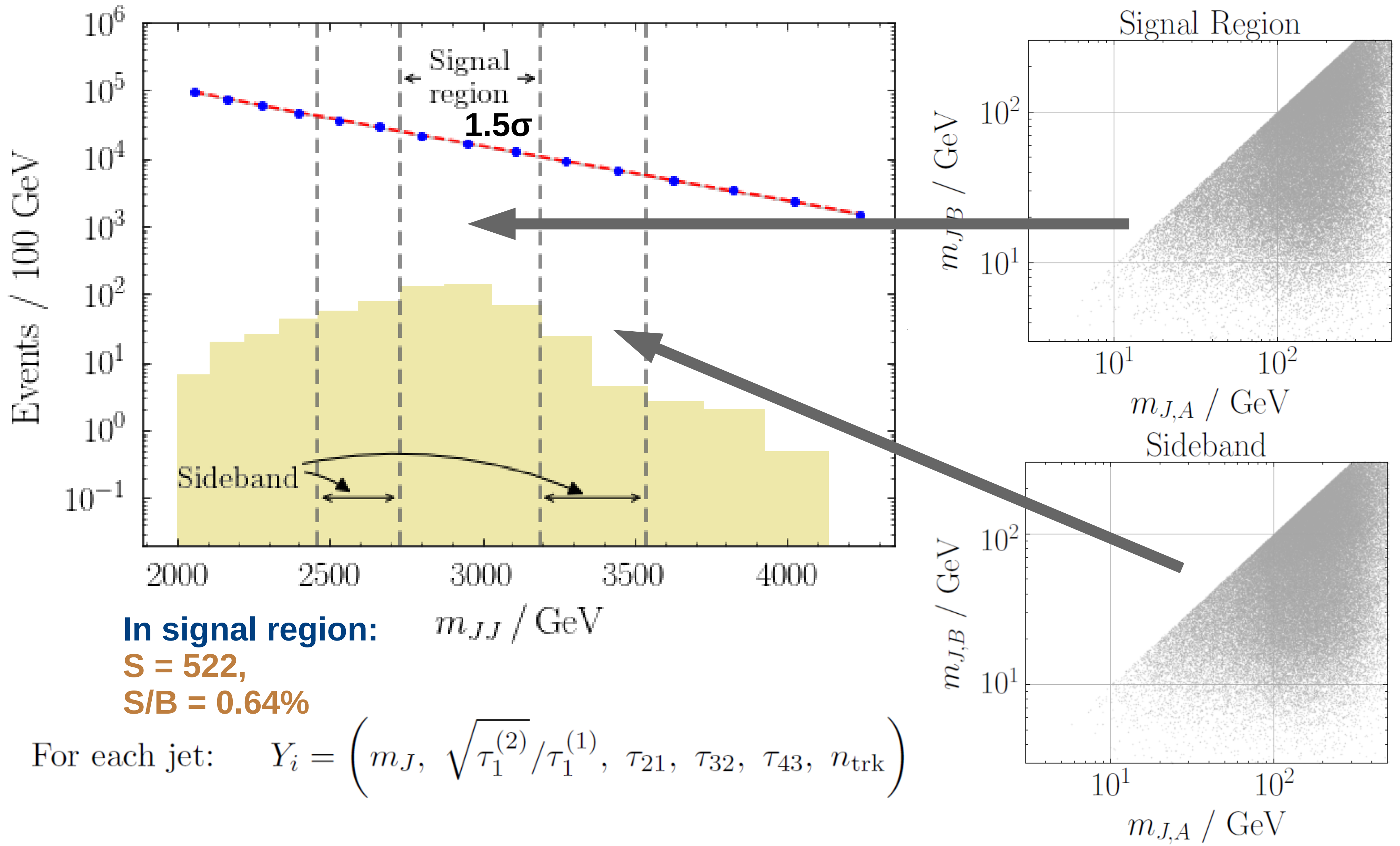


$$\frac{dN}{dm_{JJ}} = p_0 \frac{(1 - m_{JJ}/\sqrt{s})^{p_1}}{(m_{JJ}/\sqrt{s})^{p_2}}$$



In signal region:
S = 522,
S/B = 0.64%

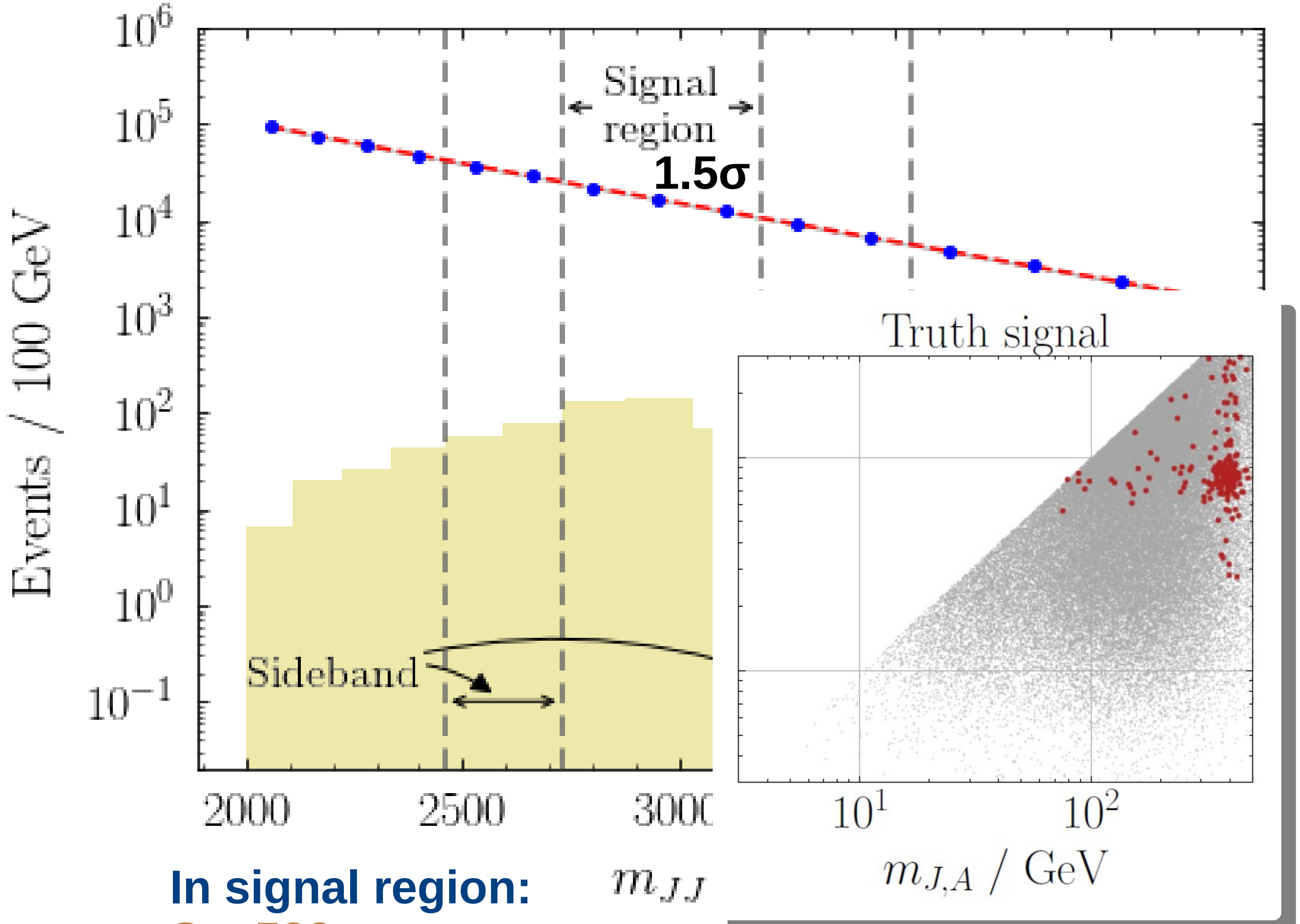
Application to Bump Hunt

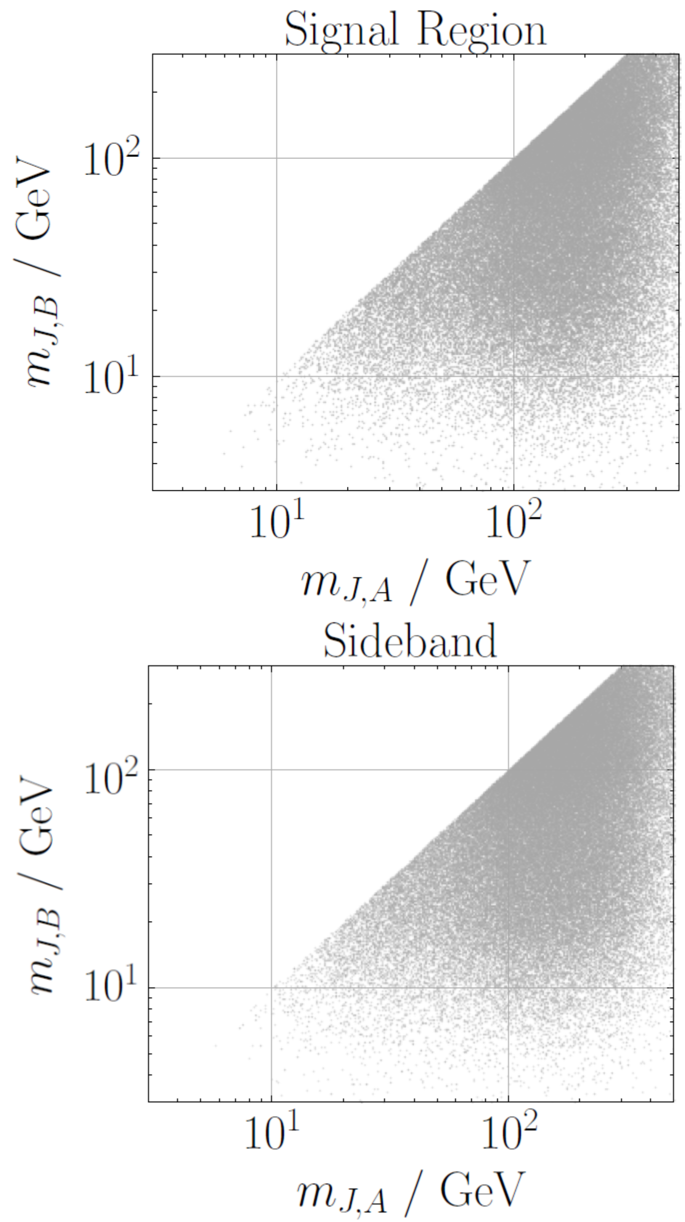


In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\text{trk}} \right)$$

Application to Bump Hunt

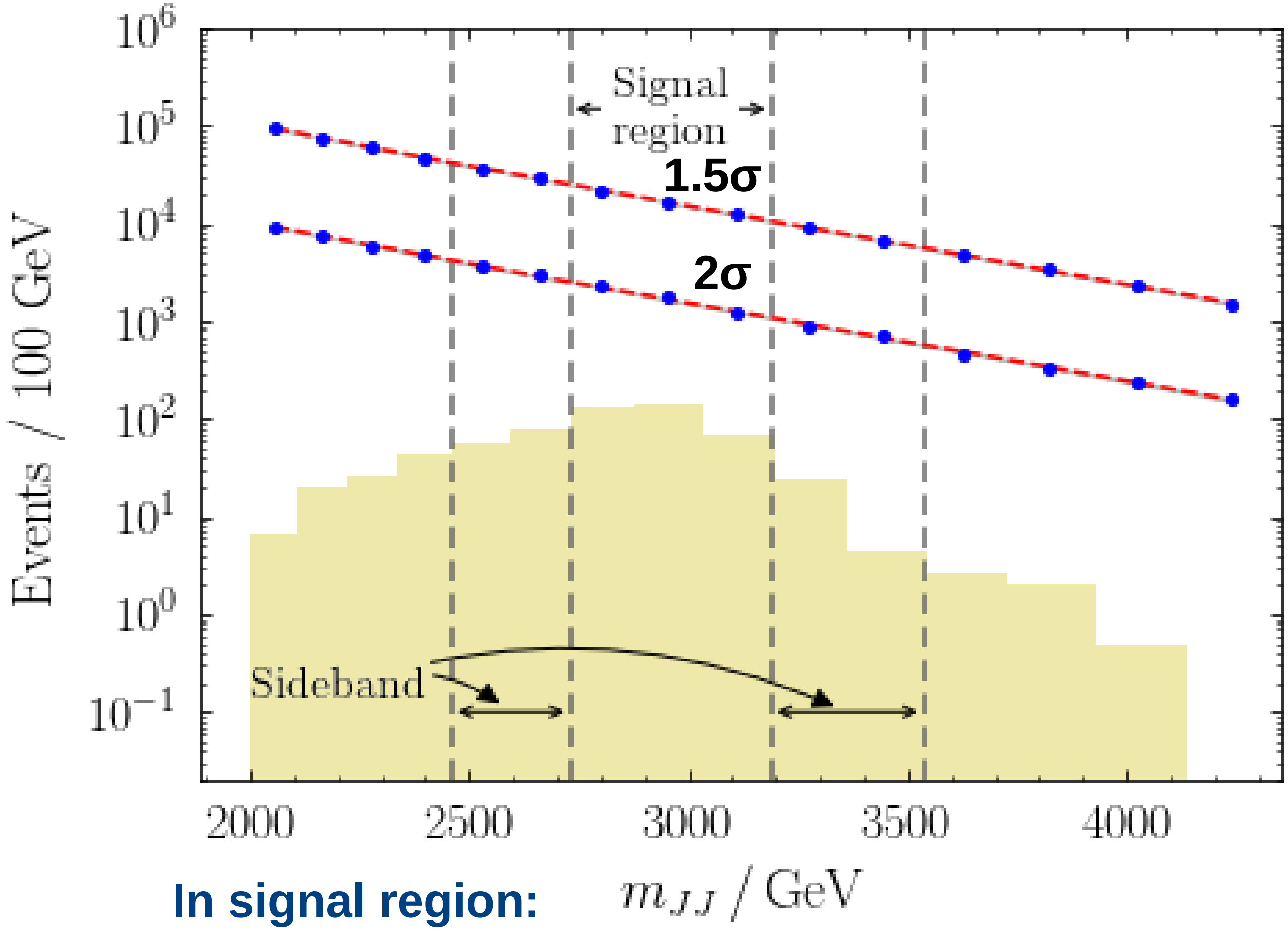


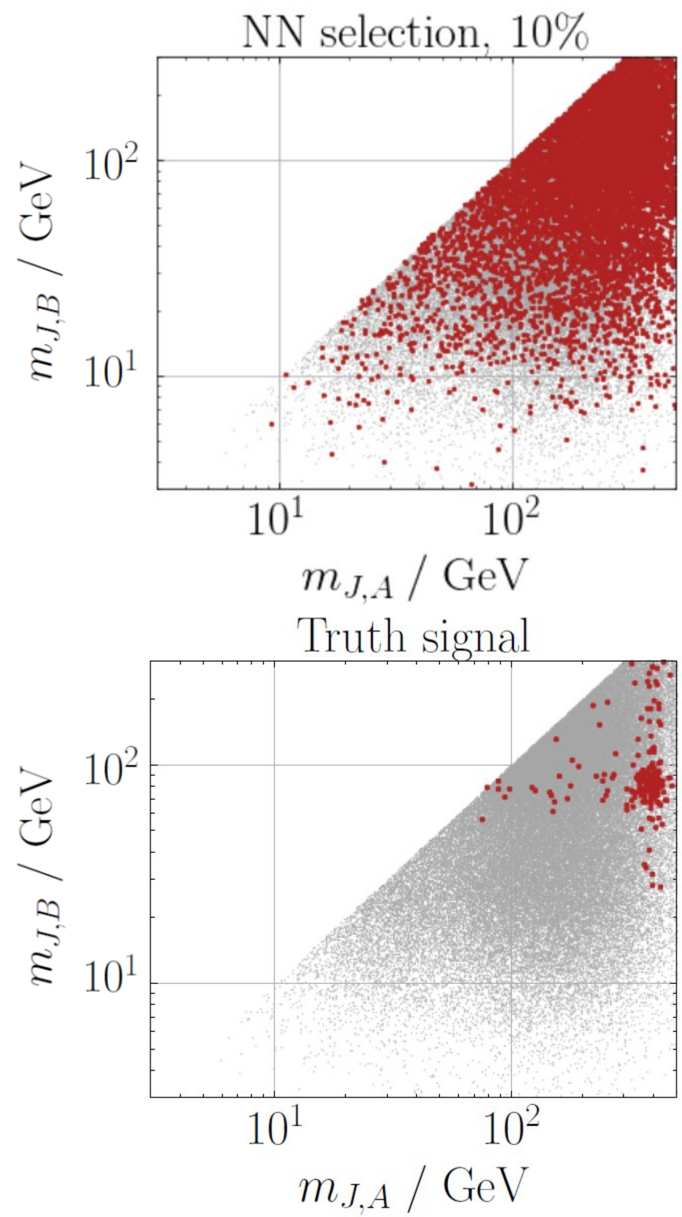


In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\rm trk} \right)$$

Application to Bump Hunt

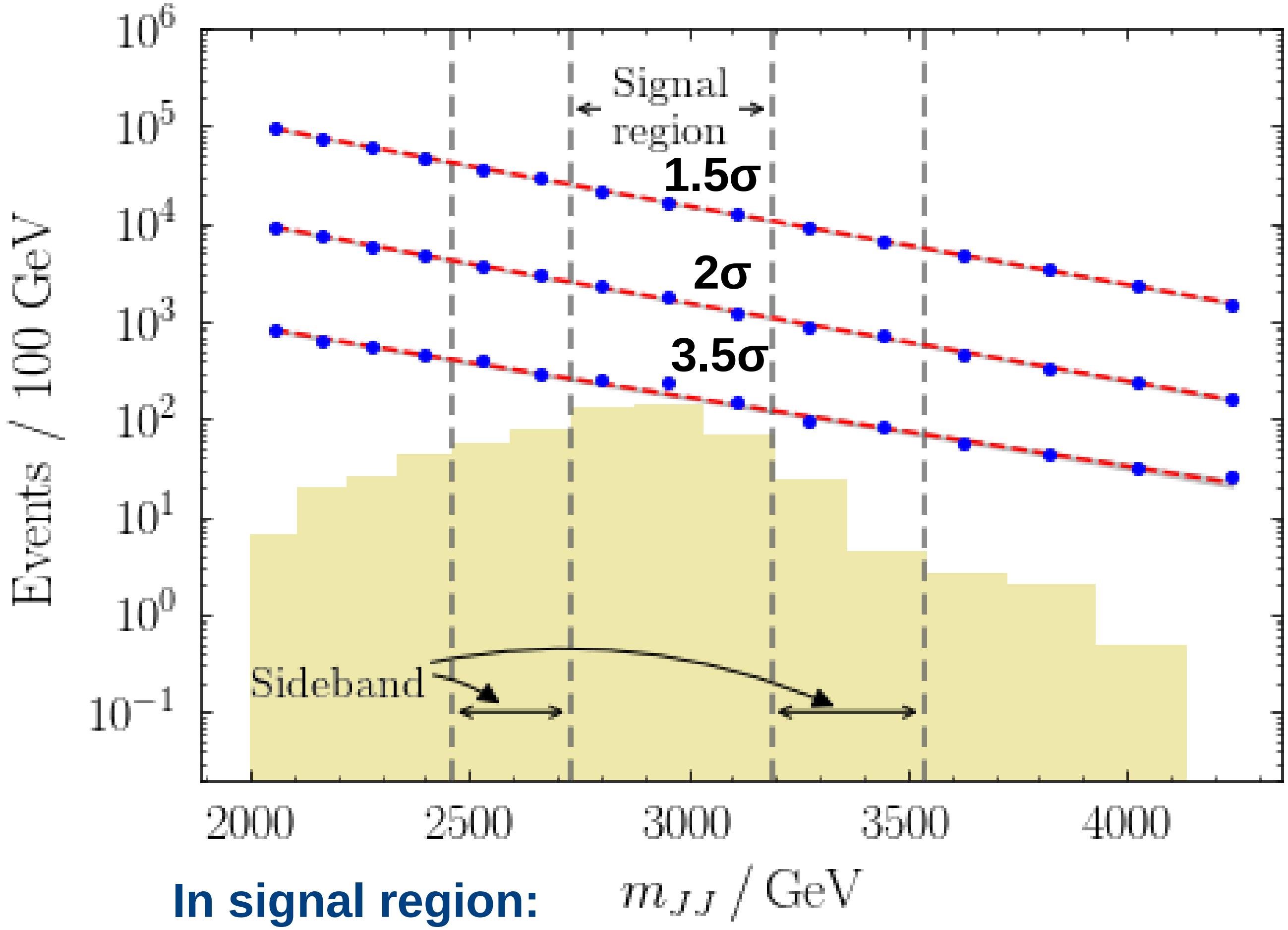


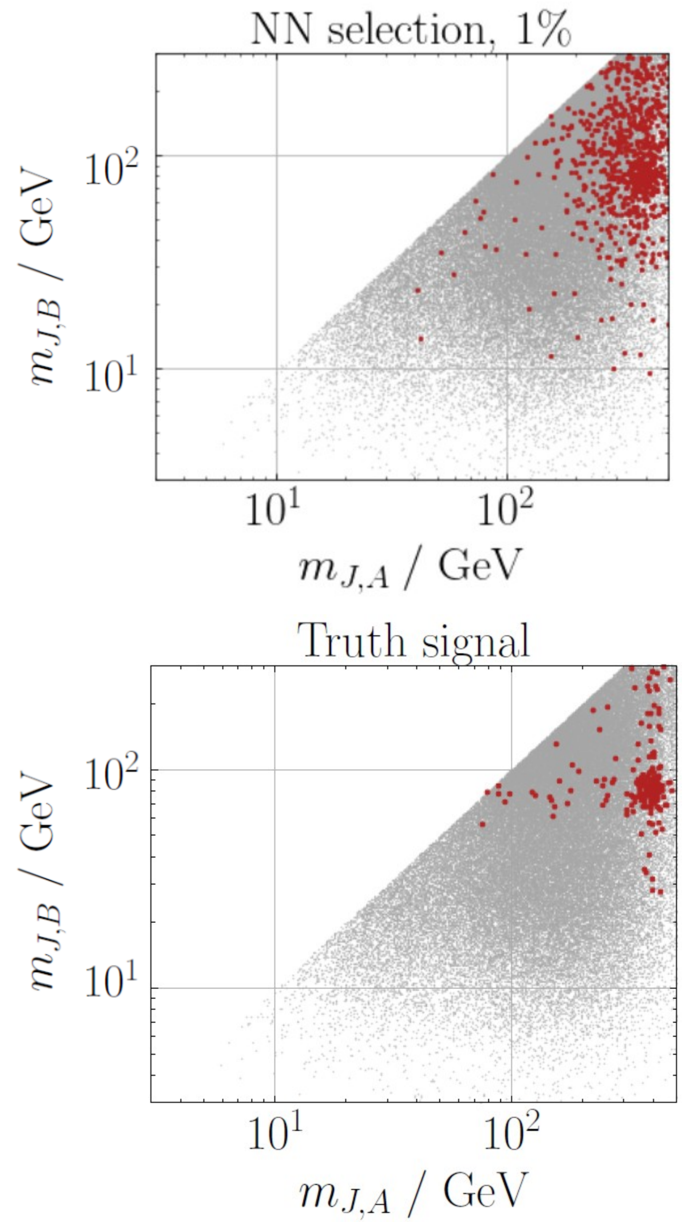


In signal region:
S = 522,
S/B = 0.64%

For each jet: $Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}}\right)$

Application to Bump Hunt

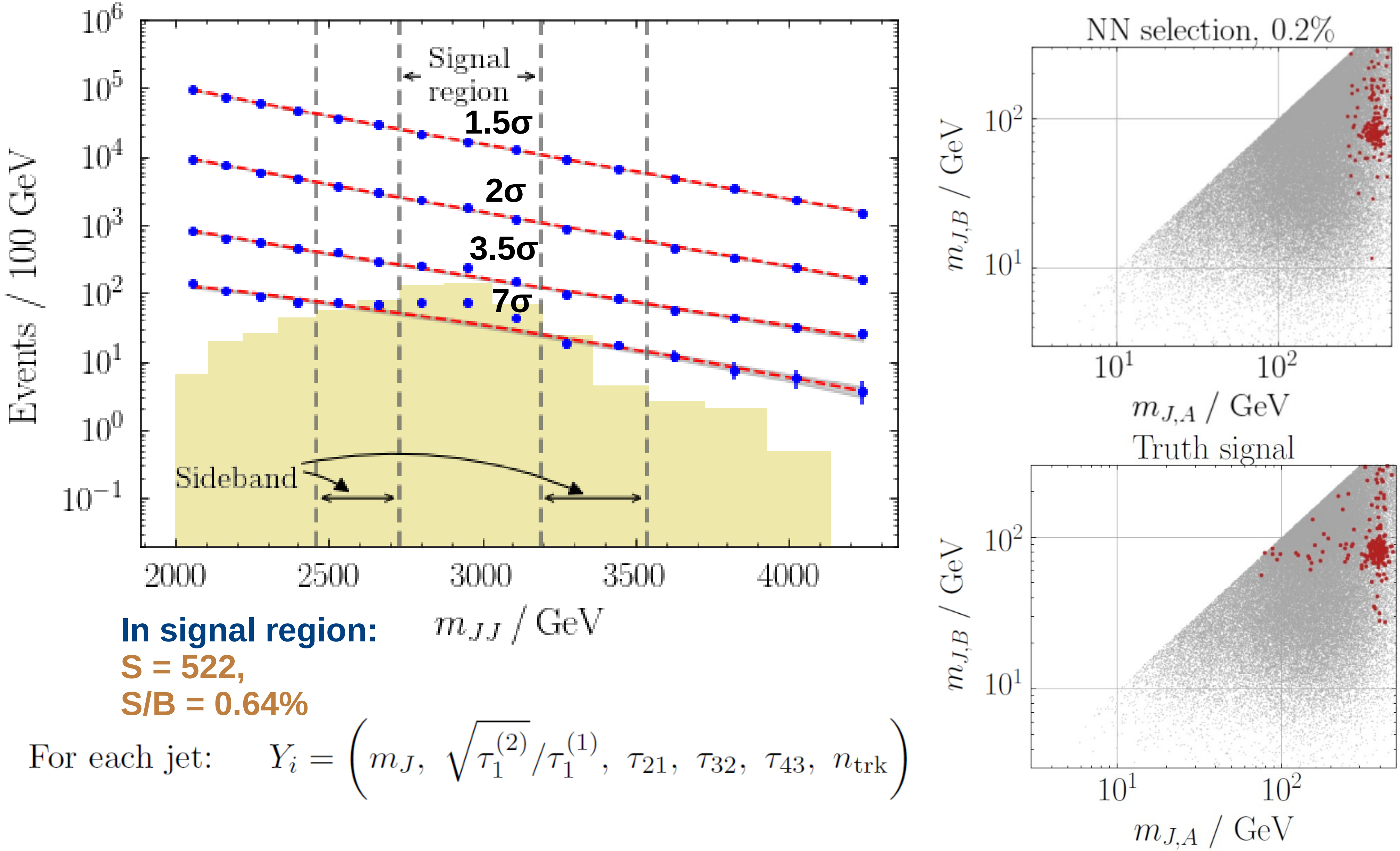




In signal region:
S = 522,
S/B = 0.64%

For each jet: $Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}}\right)$

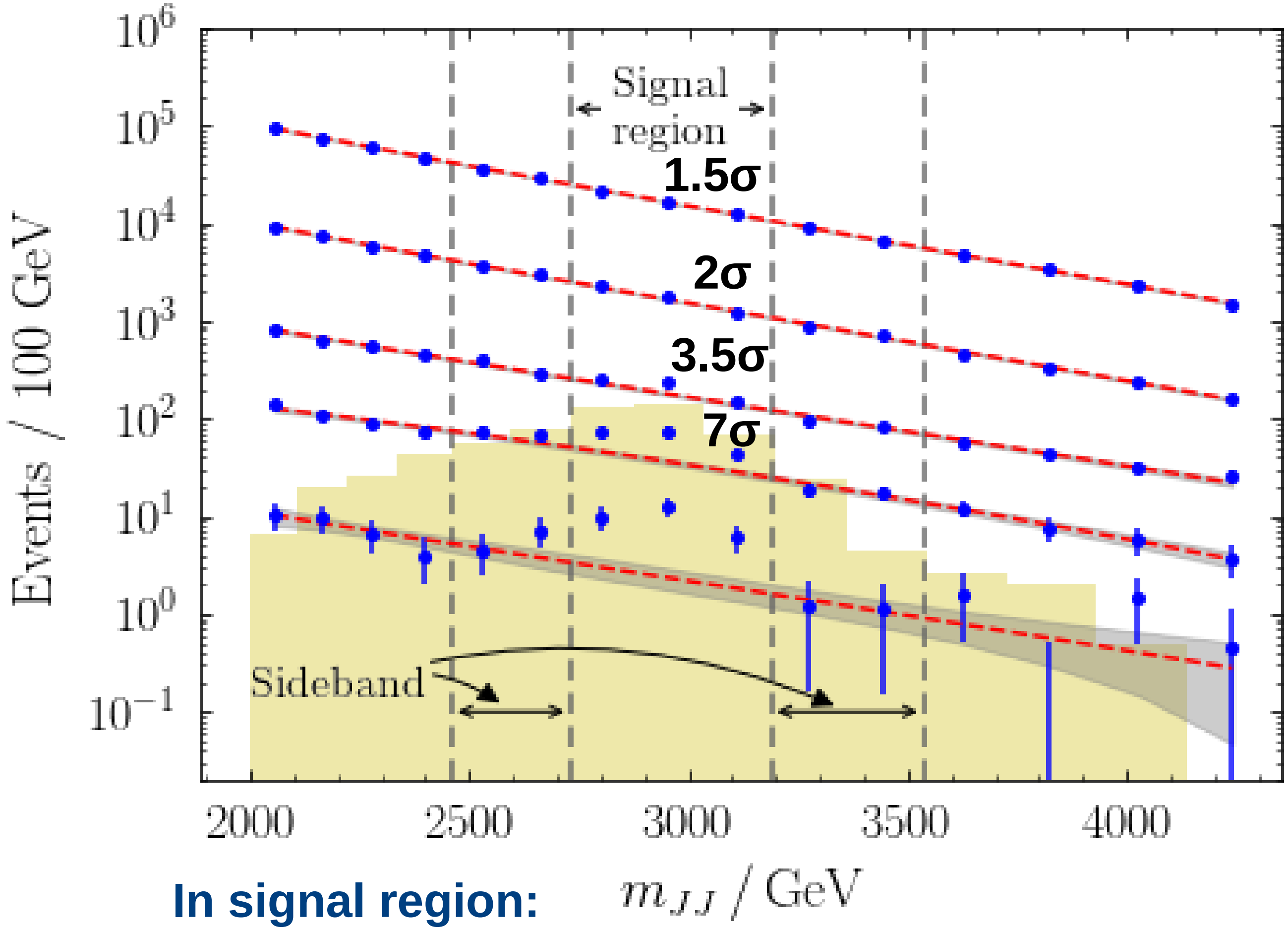
Application to Bump Hunt

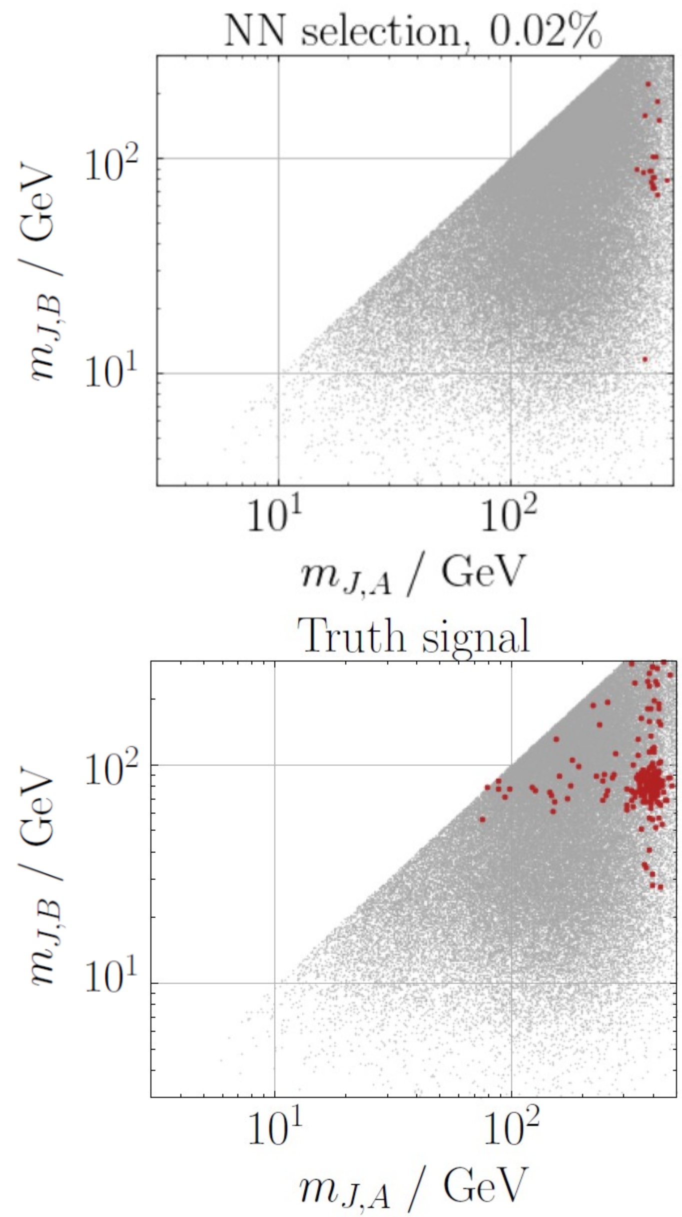


In signal region:
S = 522,
S/B = 0.64%

For each jet: $$Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$$

Application to Bump Hunt



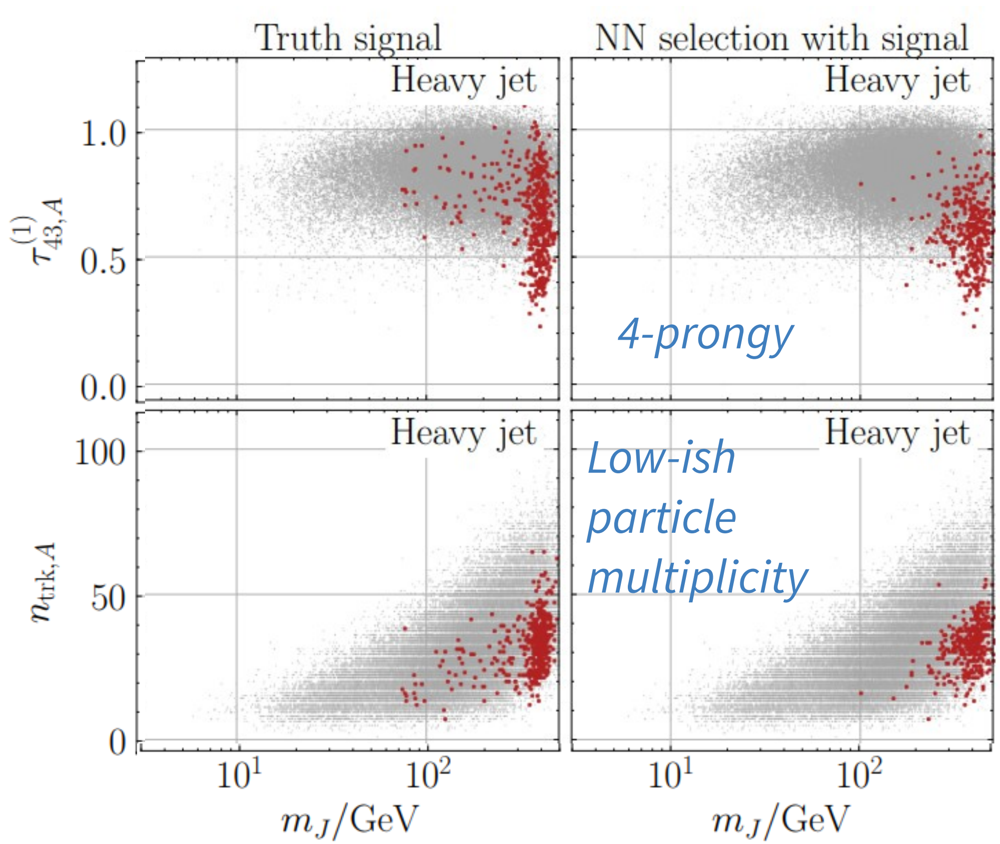


In signal region:
S = 522,
S/B = 0.64%

For each jet: $Y_i = \left(m_J,\ \sqrt{\tau_1^{(2)}}/\tau_1^{(1)},\ \tau_{21},\ \tau_{32},\ \tau_{43},\ n_{\mathrm{trk}} \right)$

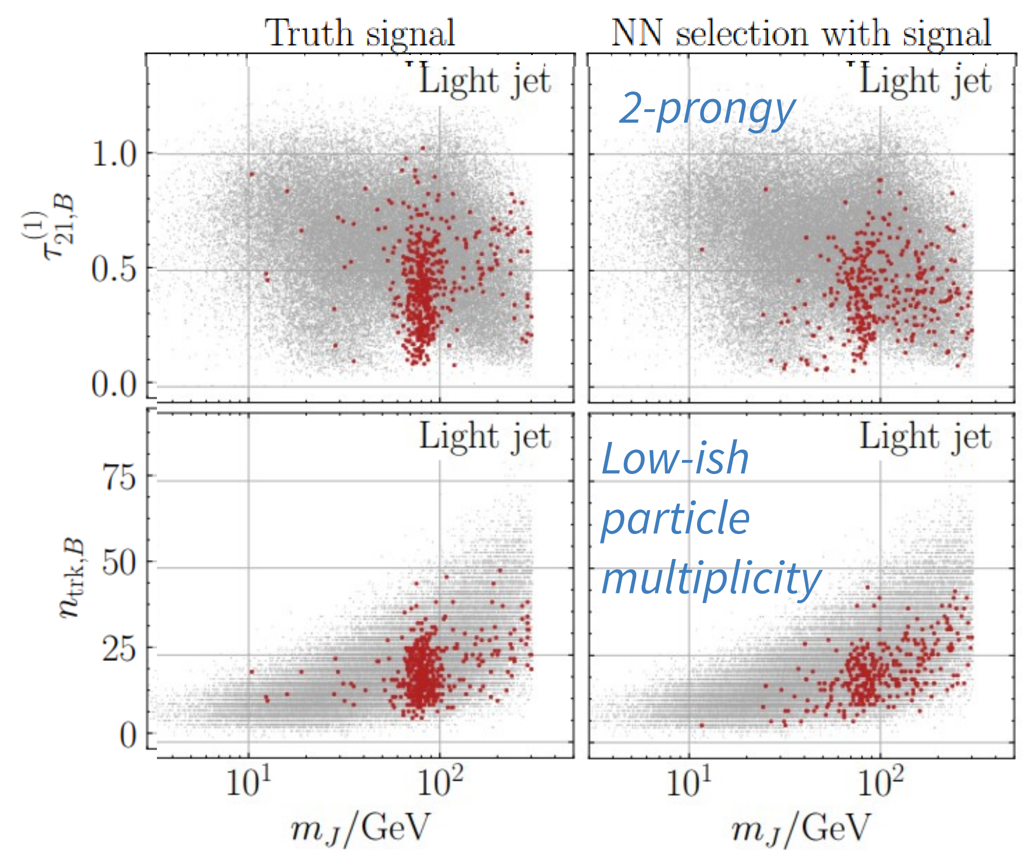
What has the machine learnt?

Jet 1



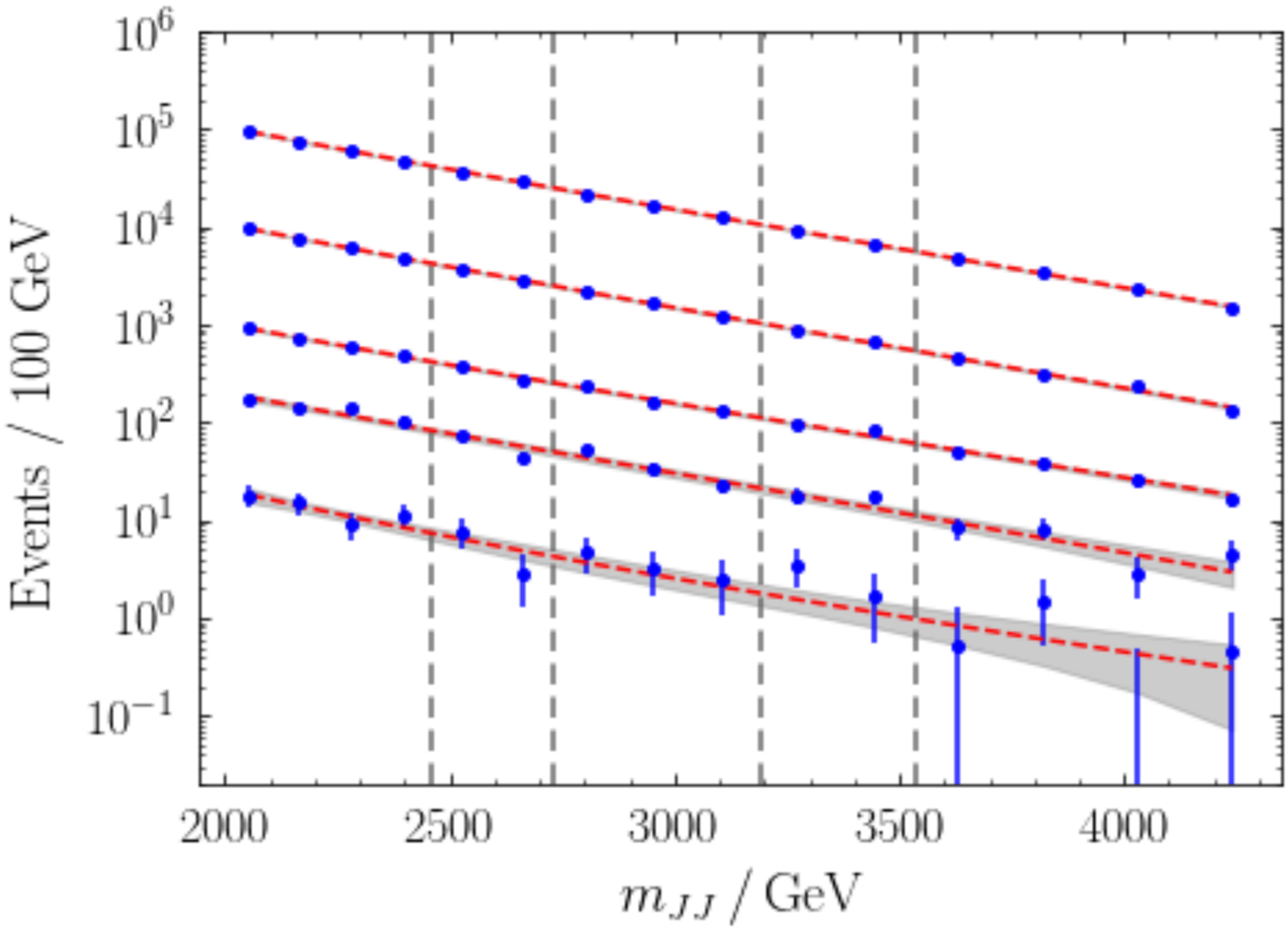
High mass

Jet 2

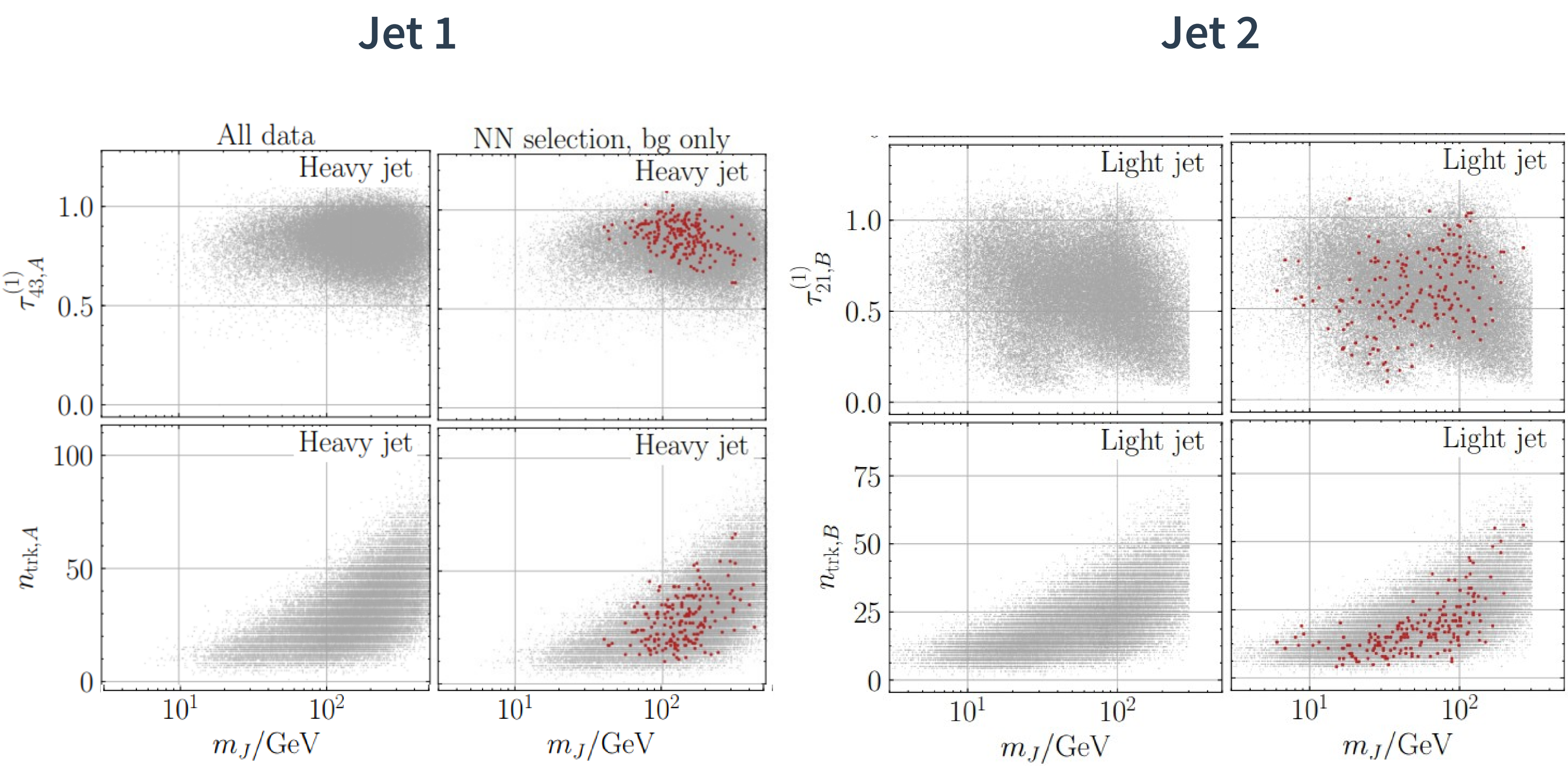


Moderate mass

No Signal → No Bump!

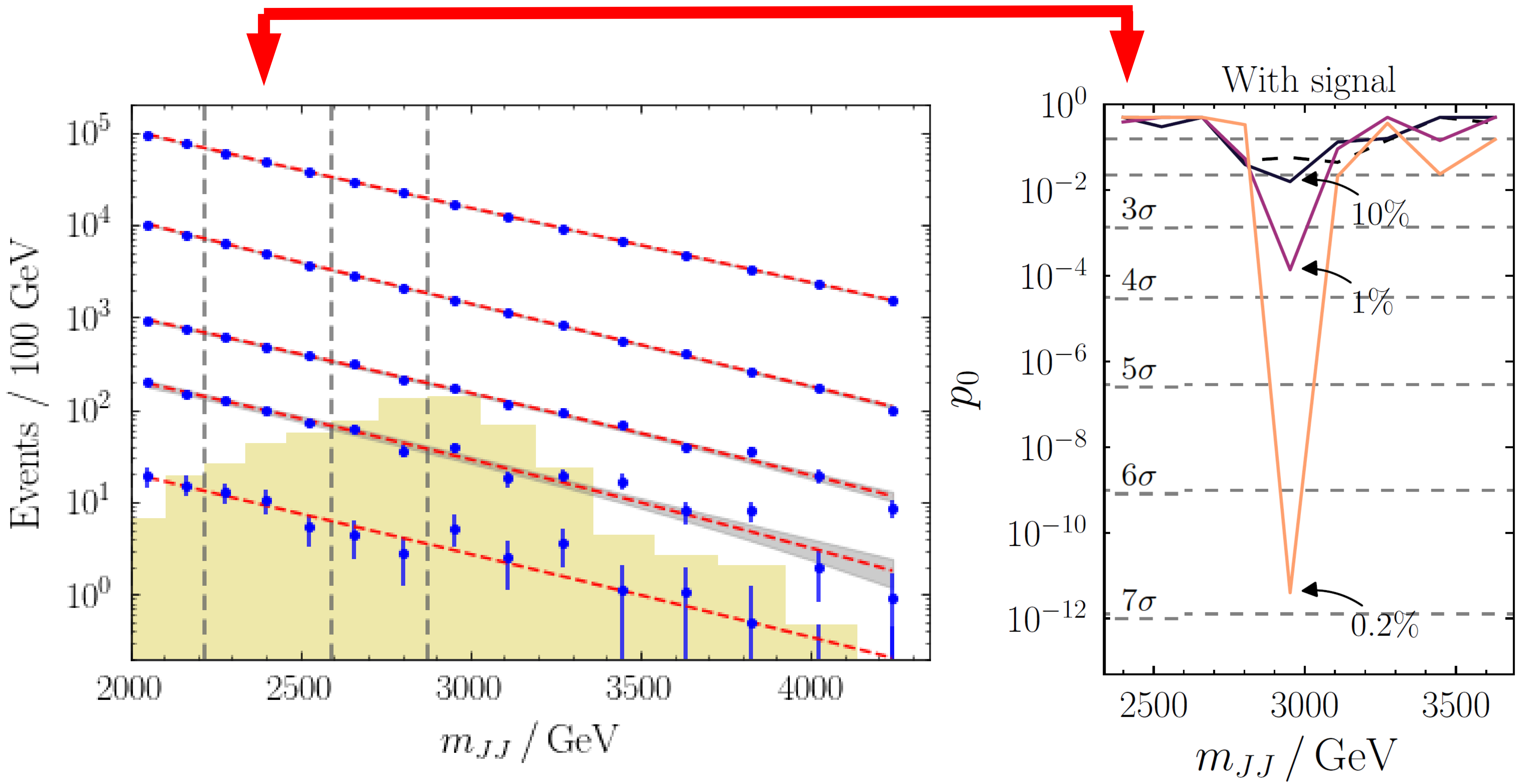


What has the machine learnt?



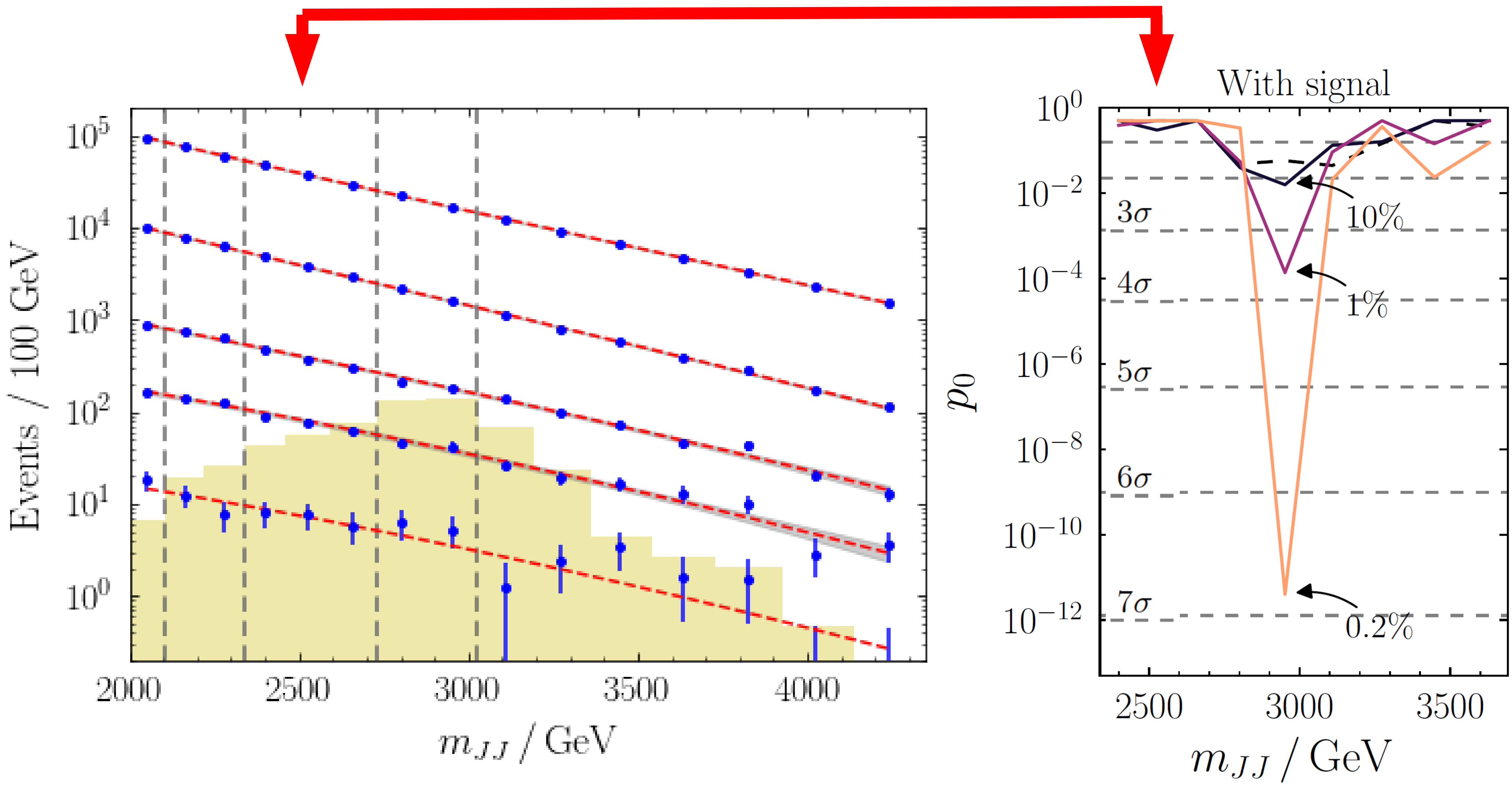
Nothing, as desired!

Mass Scan



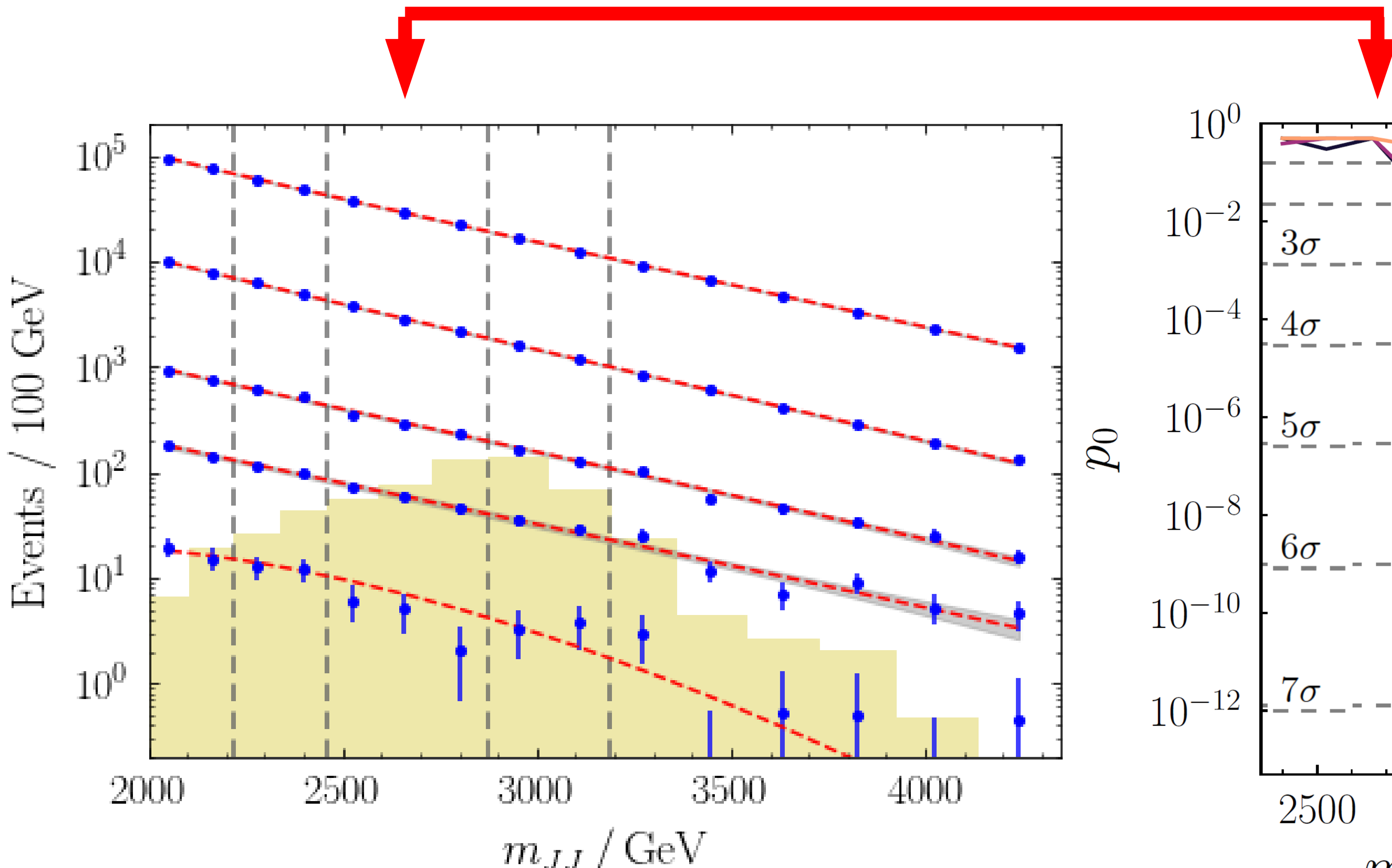
Events / 100 GeV
m_{JJ} / GeV
With signal
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
m_{JJ} / GeV

Mass Scan

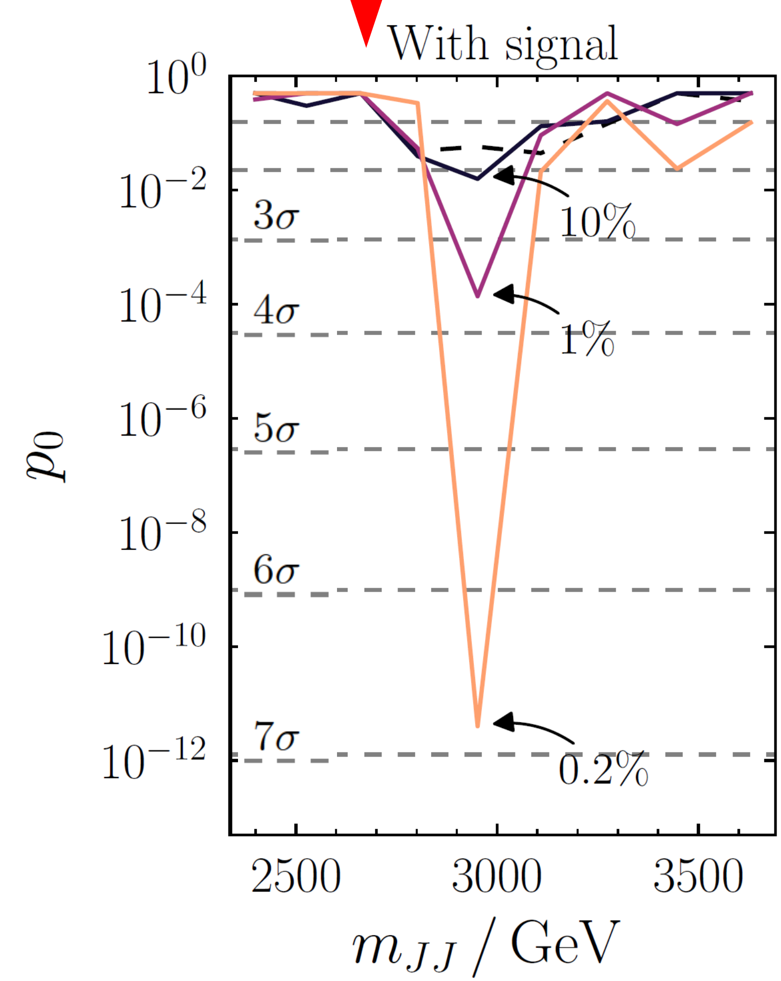


Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan

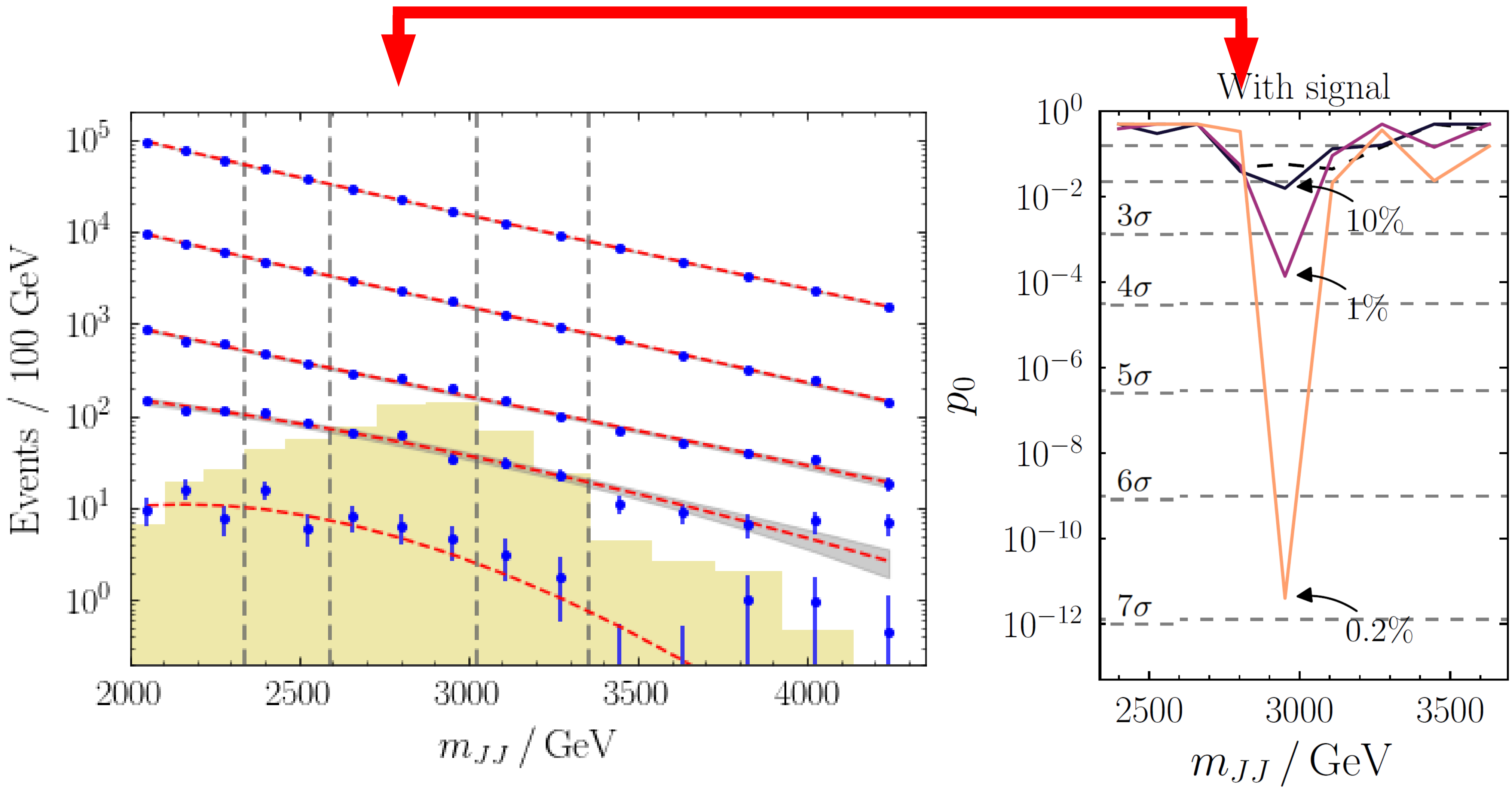


Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV



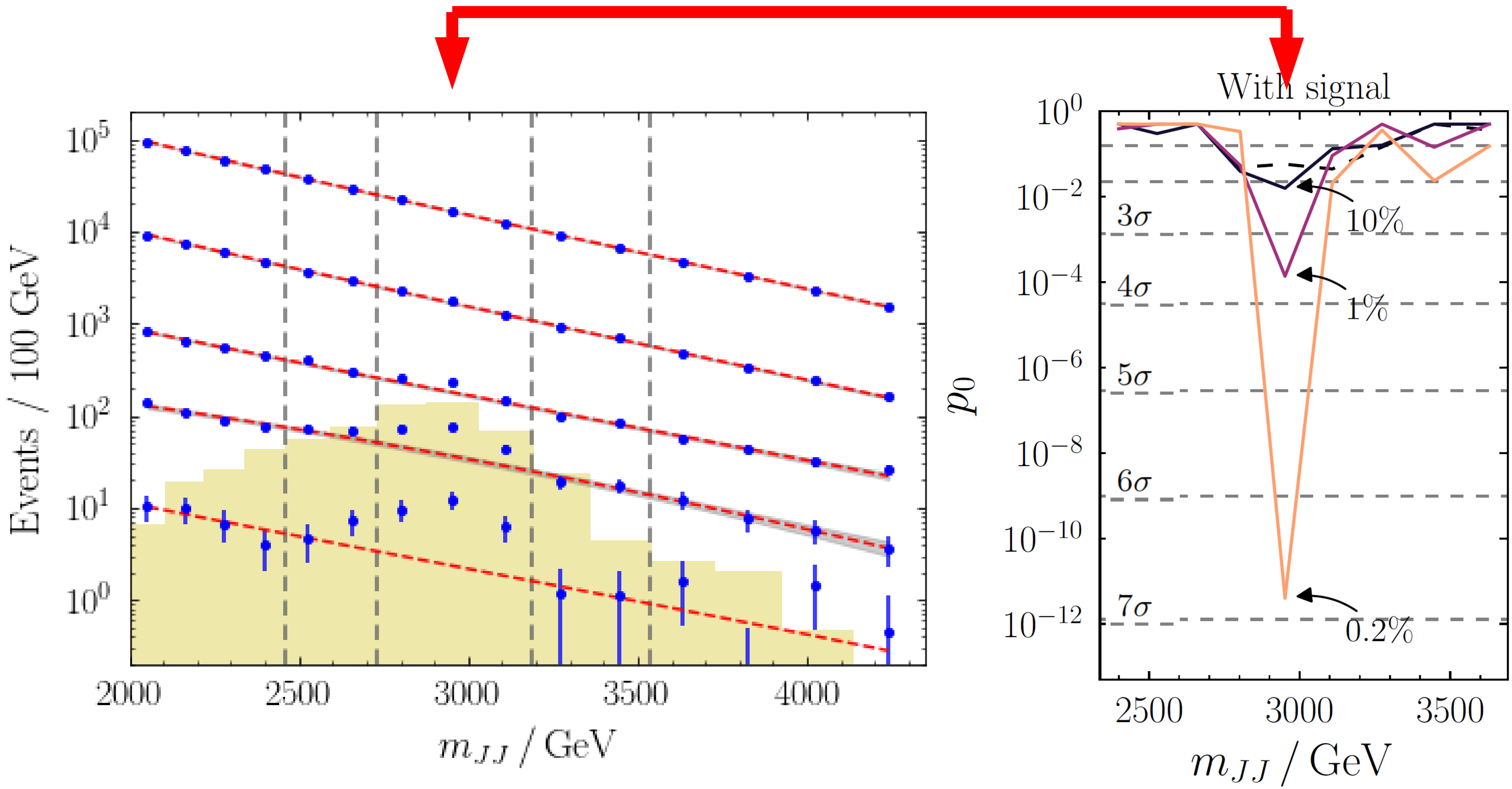
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



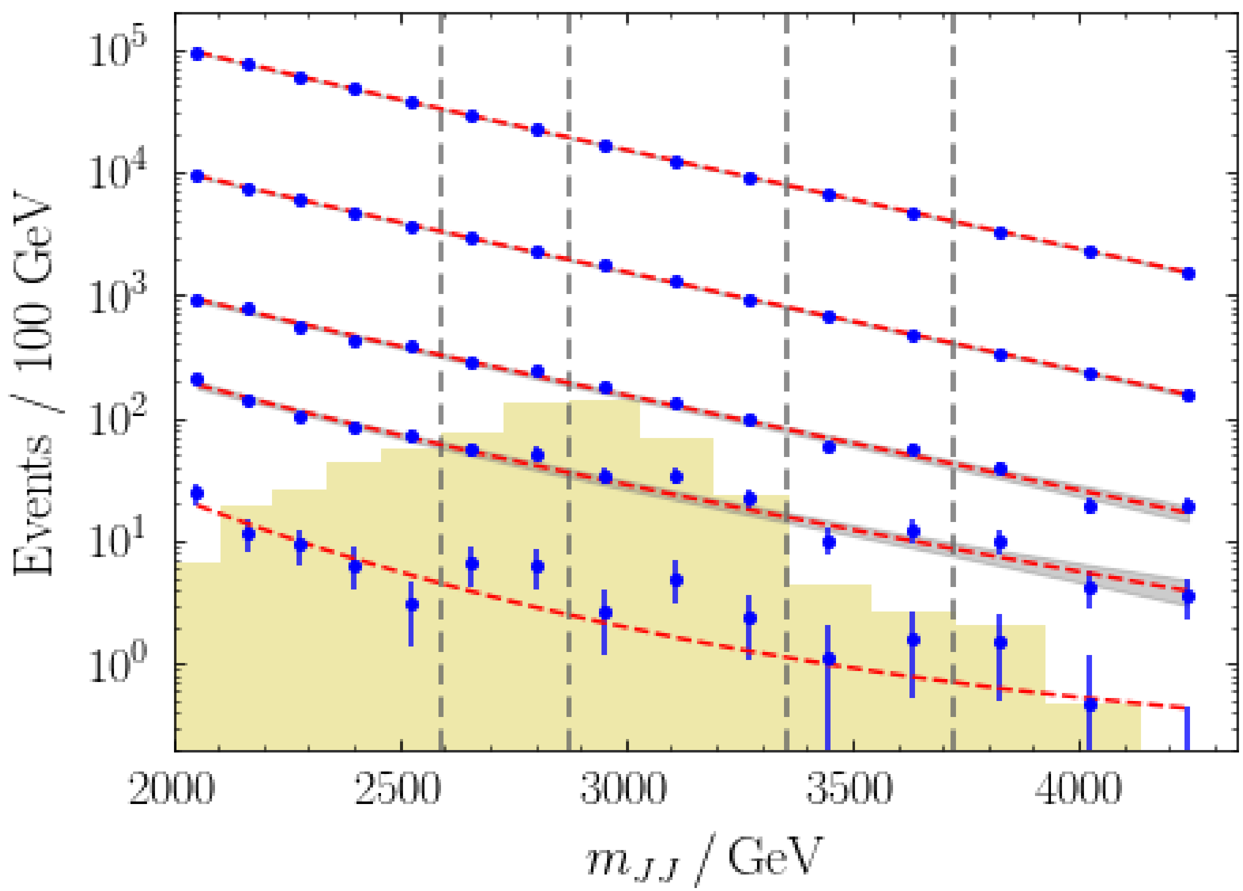
With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2000
2500
3000
3500
4000
10^5
10^4
10^3
10^2
10^1
10^0

Mass Scan

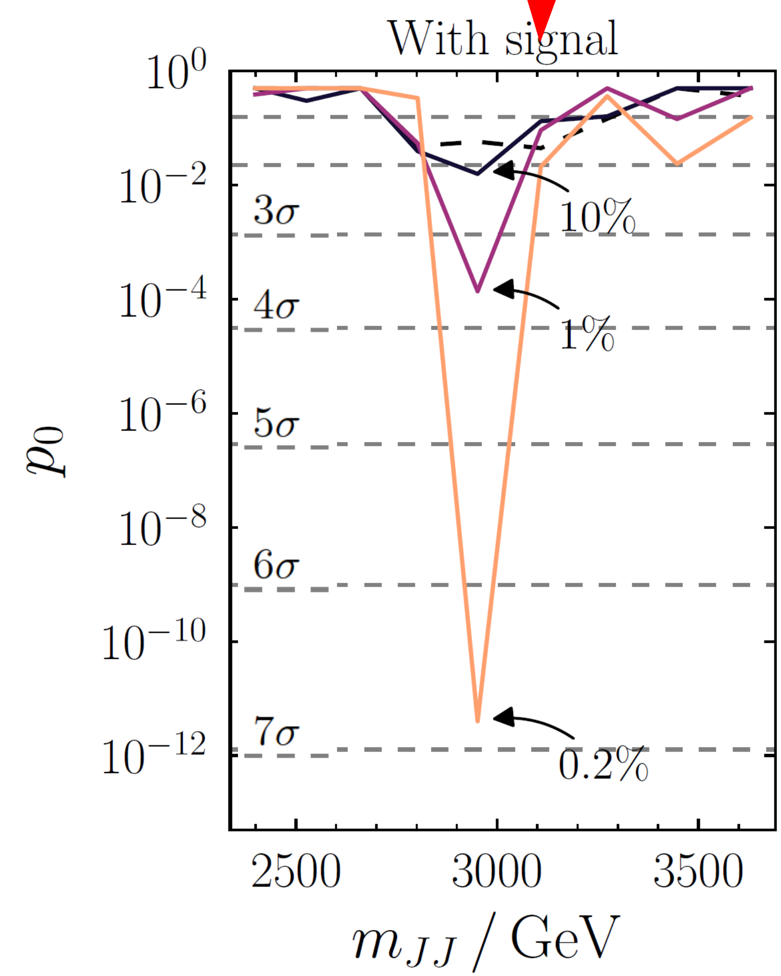


With signal
Events / 100 GeV
m_{JJ} / GeV
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%

Mass Scan

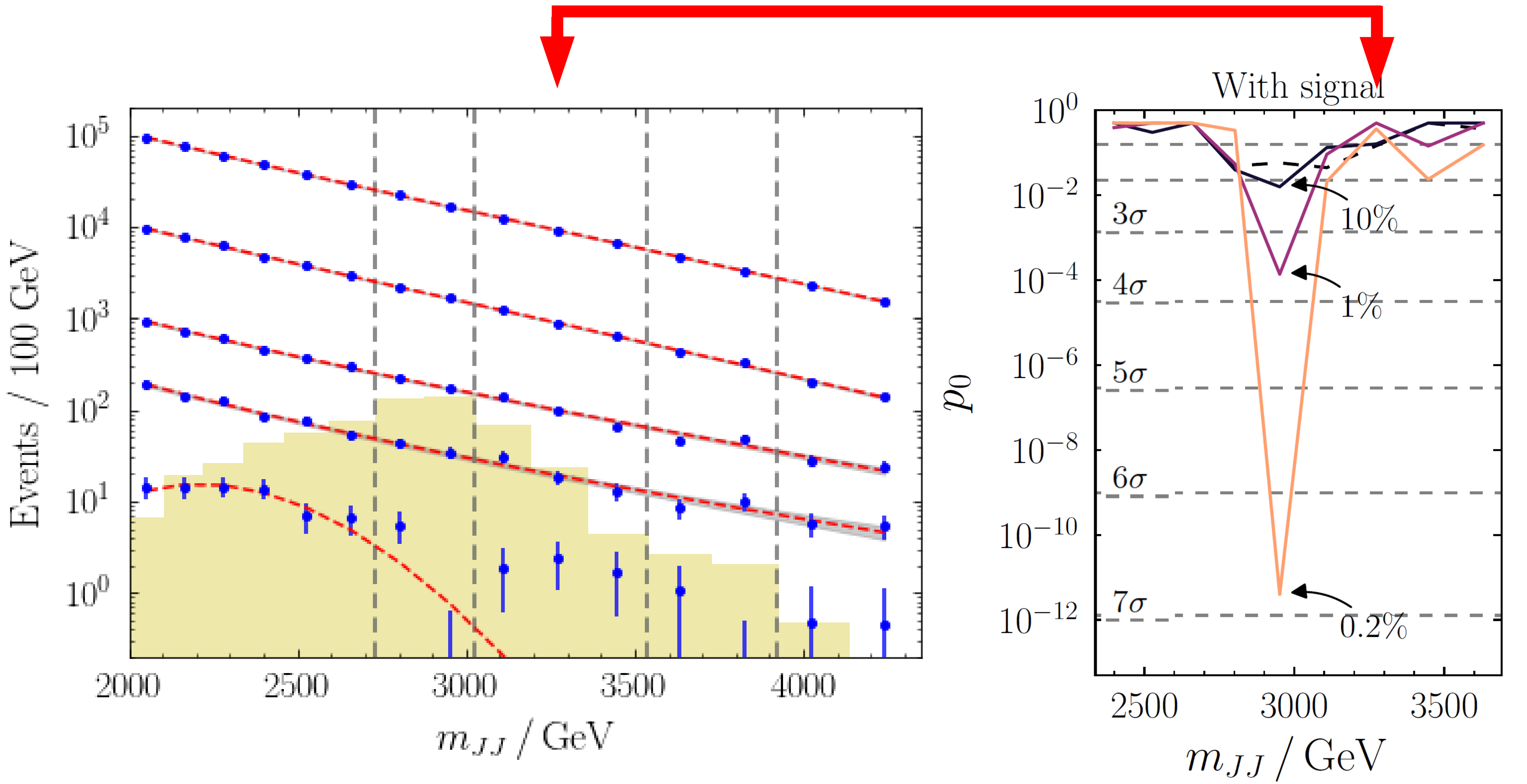


10^5
10^4
10^3
10^2
10^1
10^0
Events / 100 GeV
2000
2500
3000
3500
4000
m_{JJ} / GeV



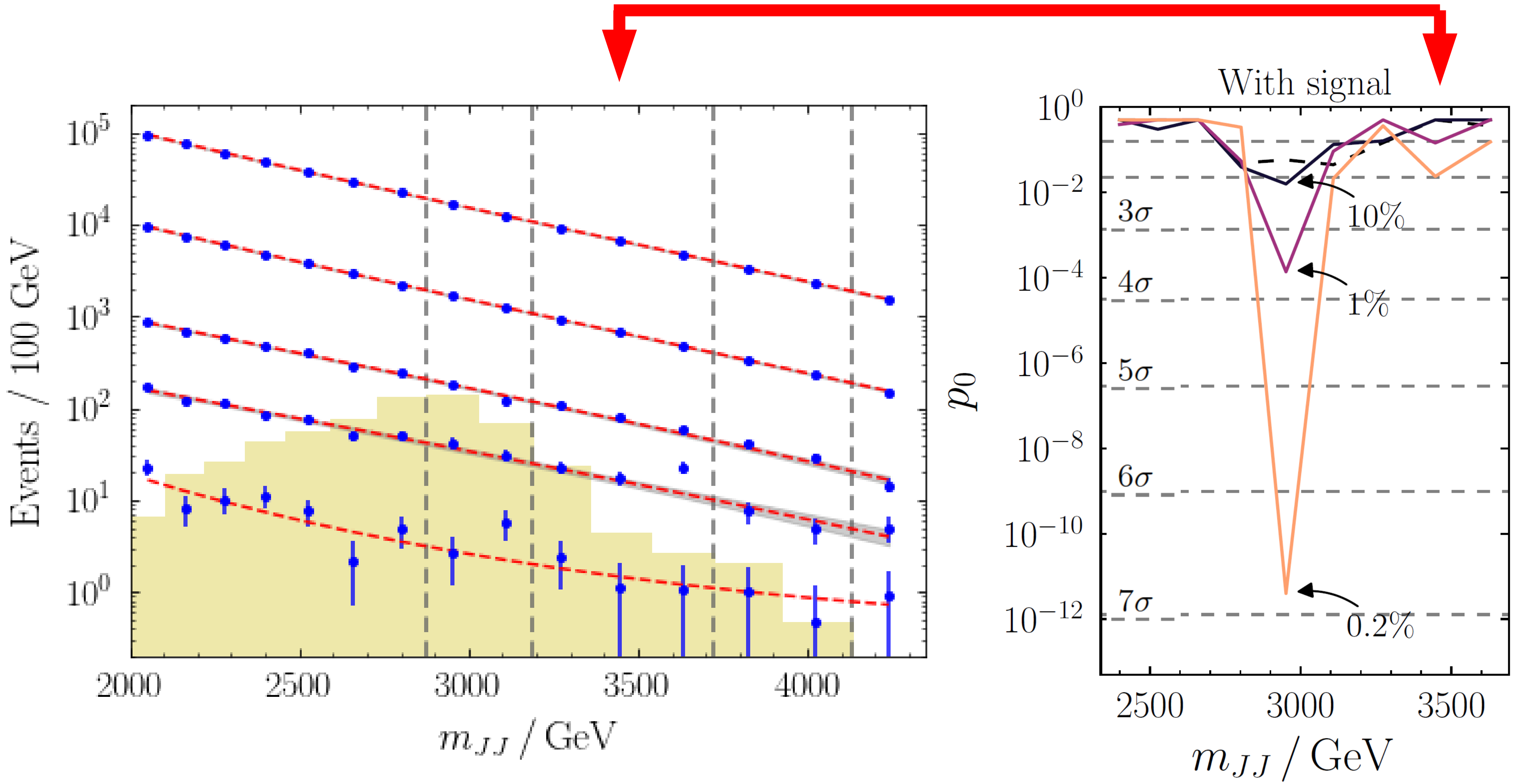
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
p_0
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



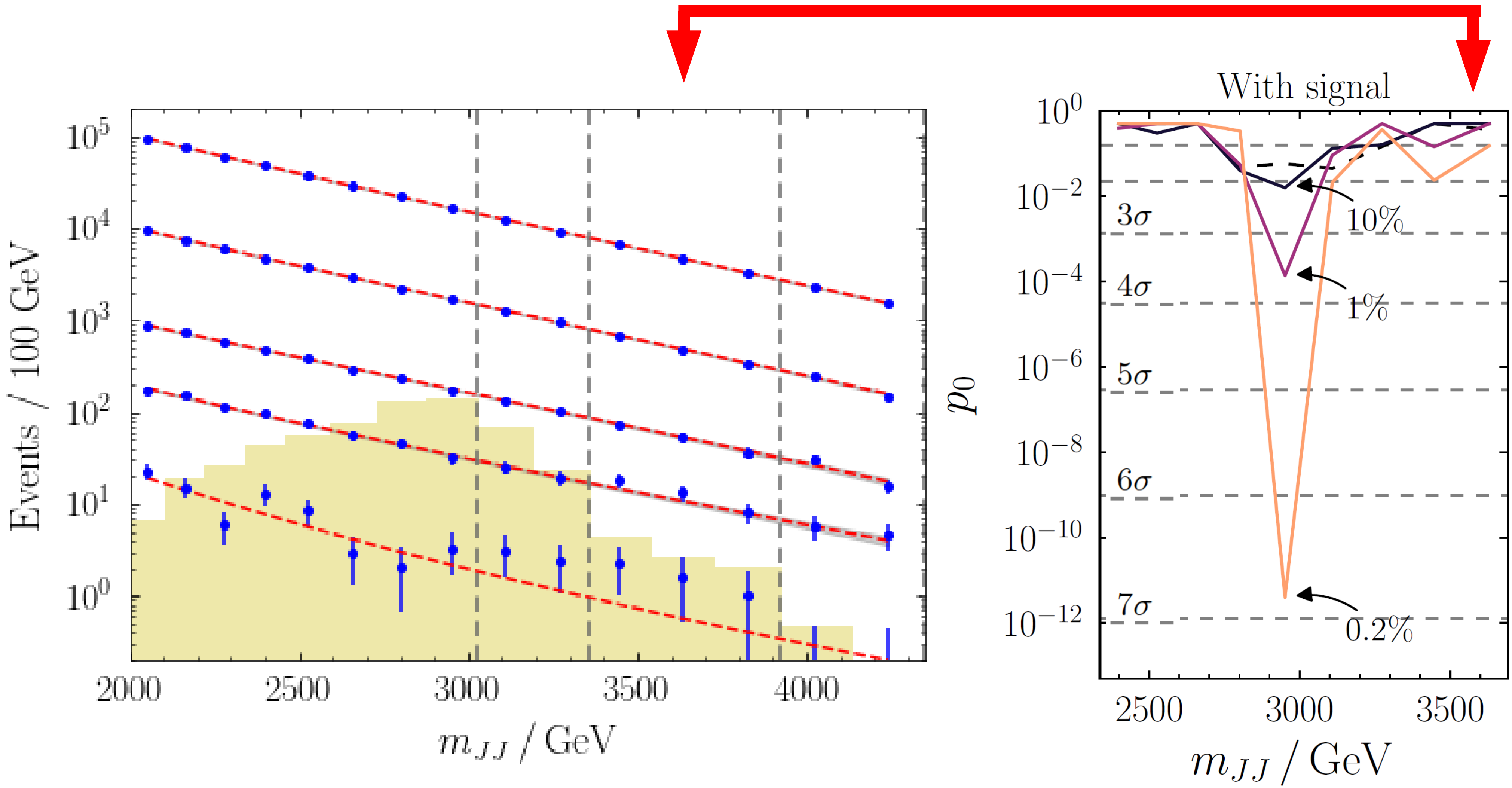
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



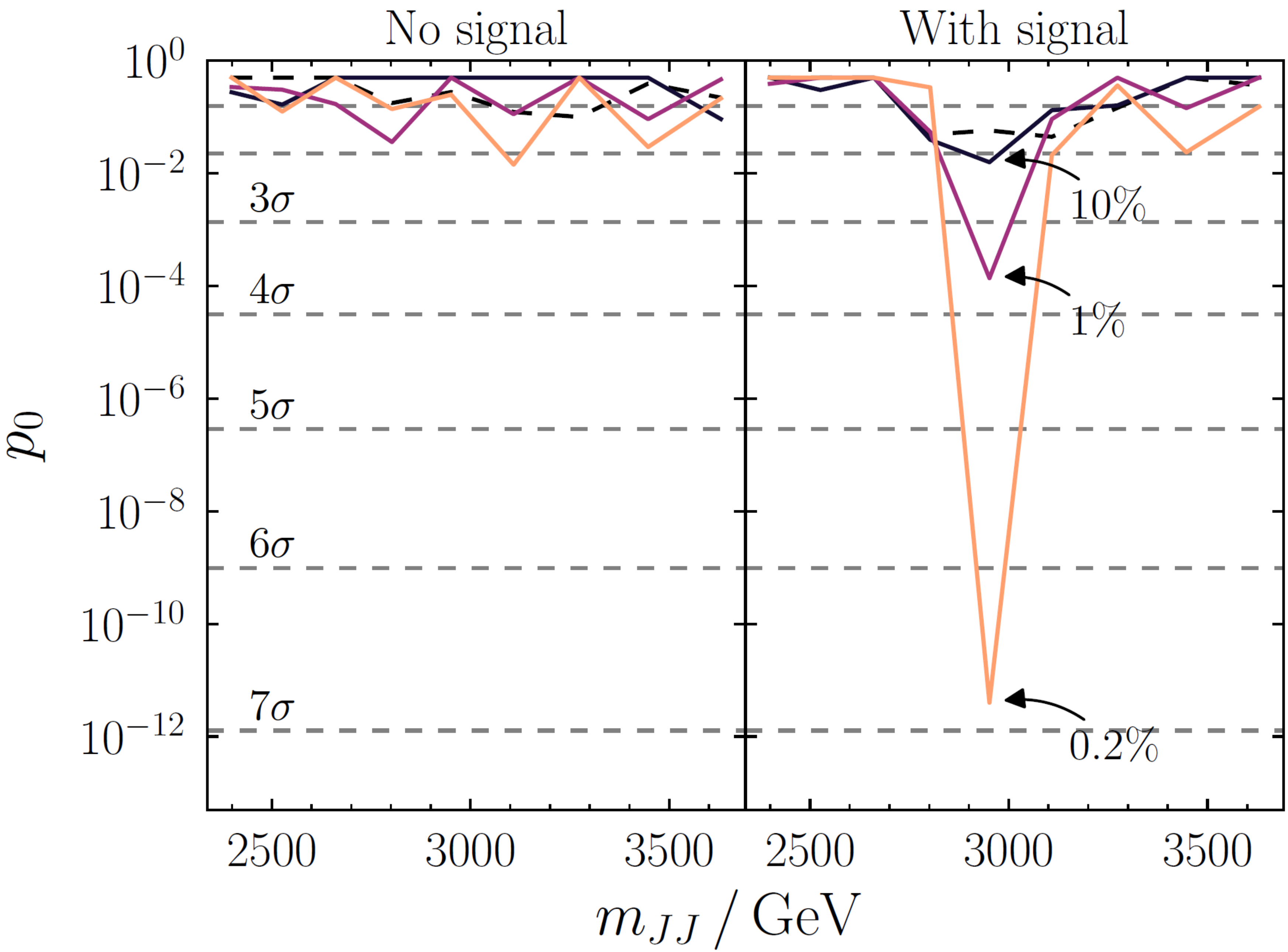
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_{JJ} / GeV
With signal
p_0
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_{JJ} / GeV

Mass Scan



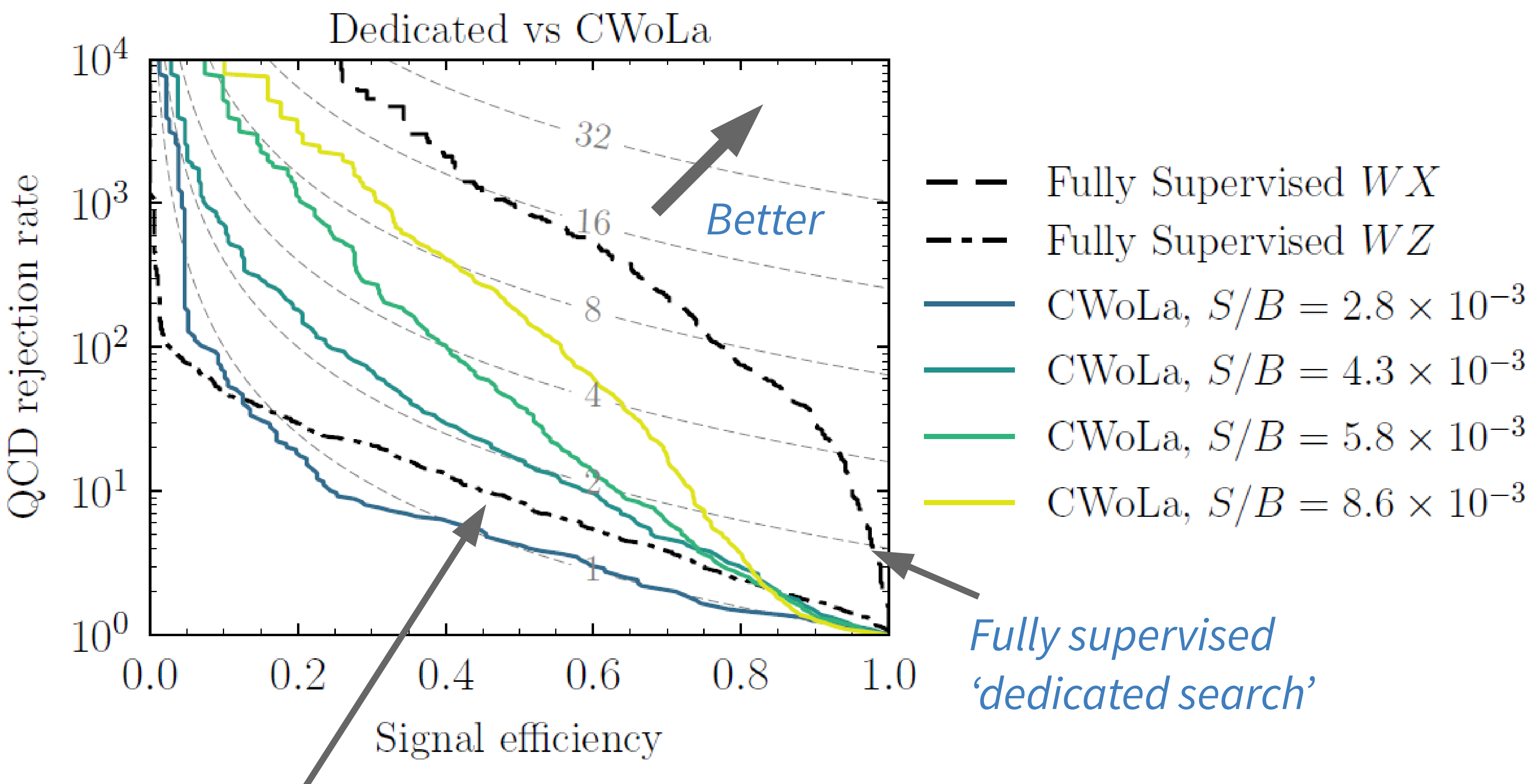
Events / 100 GeV
10^5
10^4
10^3
10^2
10^1
10^0
2000
2500
3000
3500
4000
m_JJ / GeV
With signal
p_0
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
3σ
4σ
5σ
6σ
7σ
10%
1%
0.2%
2500
3000
3500
m_JJ / GeV

Mass Scan



No signal
With signal
10^0
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
3σ
4σ
5σ
6σ
7σ
p_0
10%
1%
0.2%
2500
3000
3500
2500
3000
3500
m_{JJ} / GeV

Performance Comparison



Fully supervised 'dedicated search'

Fully supervised, wrong model.

General CWoLa Hunting

- **Need some variable X (e.g. m_JJ) in which bg is smooth and signal is localized**
- **Need some other variables {Y} (e.g. jet substructure) which may provide discriminating power which may be a-priori unknown.**
- **{Y} should not be strongly correlated with X over the X-width of the signal.**
- **Or alternatively, if correlated, there may be a way to decorrelate (e.g. if we can predict or measure the correlation, that can be subtracted away to create new uncorrelated variables).**
- **Can we use low level inputs rather than expert variables?**
  - Difficult to decorrelate auxiliary variables from resonance variable, but there are ways.
  - Pessimist: Only O(100) signal events → not enough to train with.
  - But can't know until someone tries it!

Other work: Autoencoders

[1808.08992] M. Farina, Y. Nakai, D. Shih

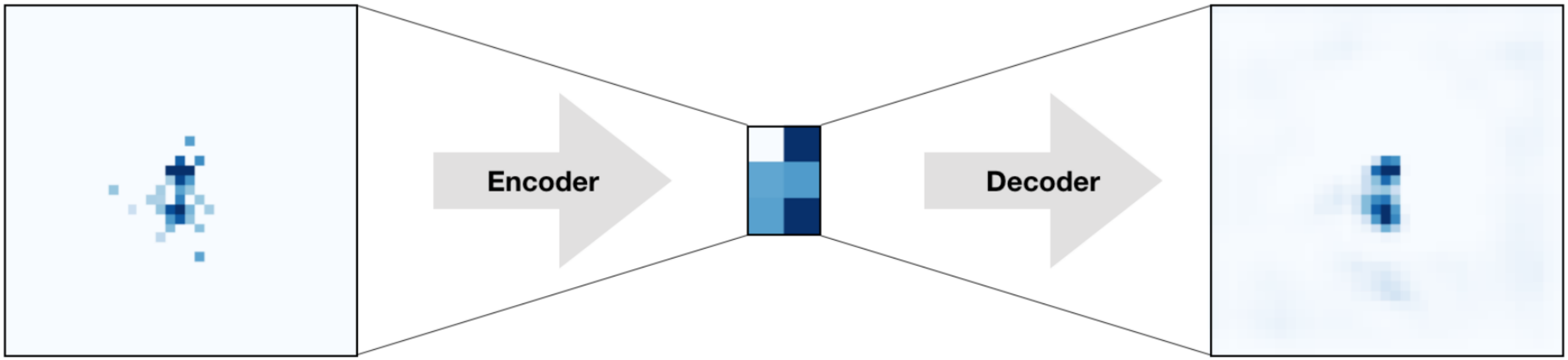


Figure 1: The schematic diagram of an autoencoder. The input is mapped into a low(er) dimensional representation, in this case 6-dim, and then decoded.

Train only on ‘background’ (no need for signal training)

Can reconstruct typical QCD background jets well, but atypical jets poorly.

→ Classify as ‘signal-like’ jets with poor reconstruction loss.

Advantage: no need for signal events for training.

Disadvantage: Can’t make use of *specific* signal characteristics for selection

Background-only training vs signal/sideband:

Background-only

Tagger performance does not depend on signal statistics.

Tagger can never learn the *specific* peculiar features of the signal, and so **cannot improve with greater signal rate**.

Signal / Sideband

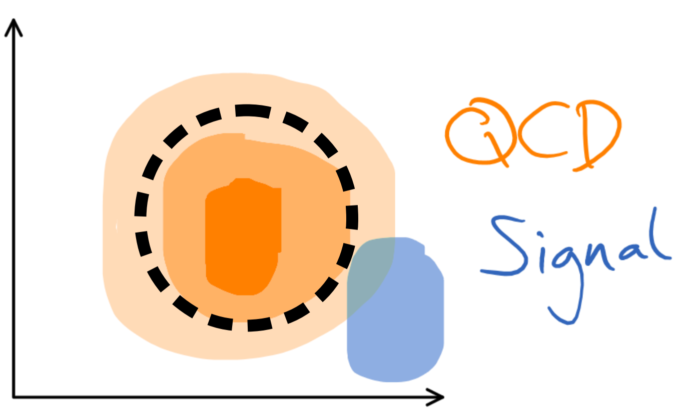
Tagger relies on there being sufficient signal statistics for training.

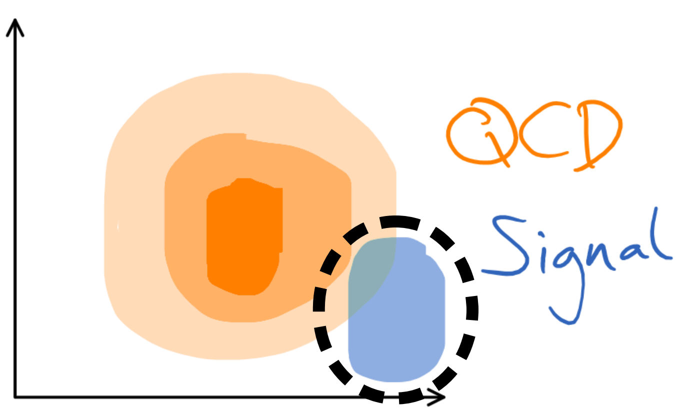
Tagger can learn the *specific* peculiar features of the signal, and so **improves with greater signal rate**, and allows for **signal characterization**.

Stronger in limit of very low signal statistics

??

Stronger in limit of very high signal statistics





Toy Statistics

$$\mathcal{L}(\mu, \theta) = \text{Poiss}(n|b + \theta + \mu)e^{-\theta^2/(2\sigma^2)}$$

$$\lambda_0 = \frac{\mathcal{L}(\mu = 0, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}$$